nature

Another nuclear power station?

Not much has happened in the nuclear power industry since Chernobyl. But now the British government has to make a decision that will determine not only British plans, but influence what others do.

FOR understandable if not particularly good reasons, the managers of nuclear power enterprises outside the Soviet Union have been circumspectly unostentatious since the Chernobyl accident. The world's two hundred or so power reactors have kept on producing electricity, and the economic benefit that goes with it, without further incident, but otherwise it has been a slack time. Even in France, which manufacturers more than 70 per cent of its electricity from uranium, there promises now to be something of a pause in the building programme. In West Germany, Chernobyl has helped to bring nearer the end of the high-temperature reactor project at Kalkar, while talk of other fresh initiatives has almost ceased.

Part of the explanation is economic; oil prices (in absolute value) are lower than for two decades, while fears of physical shortage of oil are belied by the many decisions now being made not to exploit known reserves, from Alaska to the China Sea. But a more interesting part of the explanation is the nervousness induced by the Chernobyl accident. If a crew of foolhardy reactor operators can, by wayward disregard for operating procedures, cause an accident which, tragic even in the scale of the Soviet Union, could have made many smaller countries uninhabitable, can any reactor be called "intrinsically safe"?

That is the dilemma that now confronts Mr Peter Walker, Secretary of State for Energy in the British government, who last week took delivery, at last, of the report of the government inspector, Sir Frank Layfield, who has been inquiring for the past three years into a proposal by the Central Electricity Generating Board (CEGB), the still-nationalized British electricity utility, to build a pressurized-water power station at Sizewell on the Suffolk coast. Walker, and the British government as a whole, may be forgiven for thinking it bad luck that they must decide on Sizewell now, less than eight months after Chernobyl and, even worse, no more than 18 months before the next general election in Britain.

So it is. The Layfield process has been under way, on one view, since 1972, which is when CEGB told a parliamentary committee that it would like to give the pressurized-water reactor a try. Five years later, Mr Tony Benn, one of Walker's predecessors in a previous government, said that he would sanction such a development only after a public inquiry. When in 1979 Mrs Margaret Thatcher, newly elected as prime minister, outlined a programme to order roughly one nuclear power station a year for the then decade ahead, she seemed not to feel bound by Benn's commitment. But in the event, British reactor designers have been occupied since 1981 in preparing material for Layfield. The long process could have ended at many other more convenient times. The only consolation, for Walker, is that somebody must jump first.

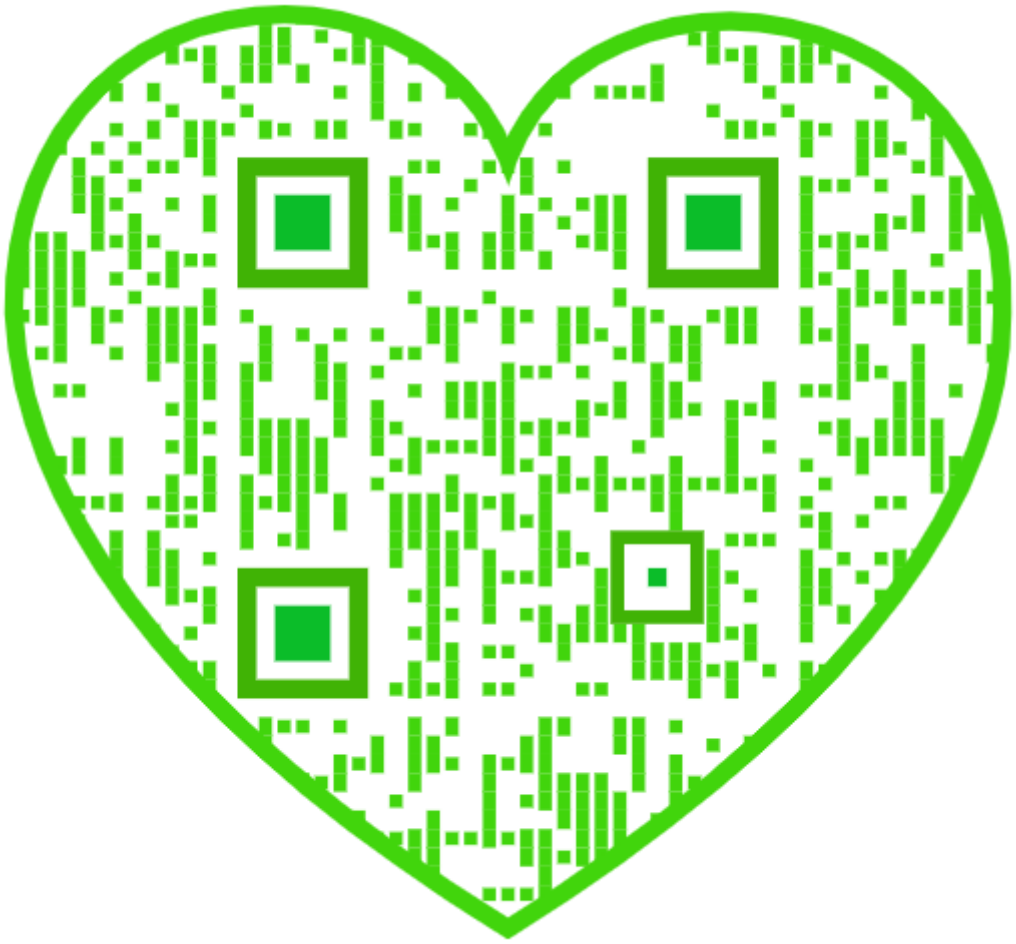
Content

In the British manner, the Layfield report has not been published. Walker would be technically within his rights to decide that it should never see the light of day, although that is so obviously a recipe for ructions as to be unthinkable. In any case, the content of the document, and its faults, are easily surmized. To the extent that safety of nuclear reactors in general and of pressurized water reactors in particular were central to the inquiry, the inspector can only say that there is no essential reason why they cannot be operated safely; to demonstrate that he was not asleep during the two long years of the public hearings, he is bound to seize on a few points in the design at which assurance could be made doubly sure by further attention to safety measures. If it had not been for Chernobyl, CEGB and its design partner Westinghouse would have sighed with relief, budgeted for the extra costs and ordered the equipment. The most obvious point on which the report (when eventually published) will be attacked is that the public hearings were completed a year ago, before Chernobyl. If the inspector has been prudent, he will have added an appendix showing that he has thought of how the inquiry might have been changed by that event. If he has not, Walker (for everybody's sake) will no doubt ask him for another piece of paper.

Scope

As public inquiries go, the Layfield exercise was unusual not only because of its length but for its scope. The intention had been to settle once and for all the issue of whether pressurized-water reactors can be safely operated in Britain, so that it might be possible to avoid the same procedure whenever another of the type seemed desirable. But, in the event, the proceedings took in all the questions relative to the good sense of generating electricity from uranium. Is all that electricity needed anyway? Will not the continued manufacture of plutonium assist the spread of nuclear weapons? And what will be done with the radioactive waste? The inquiry might have been a unique exercise in public education if it had been joined by a representative section of unofficial opinion-holders (which would have been more likely if half the time had not been spent in a remote seaside town). A wise inspector will have said that these general matters are not for him but for others, mostly for the government, the British parliament and its electors, and will then have gone on to make a few privileged unassailable observations on them. Alternatively, he may have taken a view on some general issue of policy whose delegation to him would be constitutionally improper, and may have undermined the value of his report as a whole. Either way, accepting that the British need be no less successful at operating pressurized-water reactors than people in a dozen other countries, the hard question will be left with Peter Walker.

Is nuclear power necessary, for example? That question, disarming as it may seem, is too simply asked. In the economics of energy supply, there are no categorical imperatives. Oil is an important source of energy in most industrial economies, and is a particularly convenient source of liquid fuels and of hydrocarbons in general, but economies denied access to petroleum could probably manage in other ways, but at a cost. The question to ask about the usefulness of nuclear power stations is one about the economic benefits that may be derived from them, which can be answered fully in one place only with a knowledge of what is likely to be the pattern of their use elsewhere. It goes without saying that, in present circumstances and especially when oil is cheap, it could well emerge that extra safety mea-



sures could price even the most efficient reactor designs out of the market. But Walker will also know that the fluctuating price of oil can only, in the long run, increase relative to the prices of other fuels. The need for nuclear power in Britain or anywhere is not an absolute but a matter for intricate judgement. Walker's decision will be the more convincing if he makes that plain. The Layfield report itself could turn out to be an invaluable guide to the often contradictory economics of those with opinions of the value of nuclear power — optimistic, sceptical and downright perverse.

Research

For the research community, another economic question may be even more important. Britain is not the only country whose government has decided that the technology of nuclear power is potentially too important to be left to others. Direct spending of taxpayers' money on research and development in nuclear power is probably not much short of £100 million a year, with perhaps a comparable sum creamed off electricity bills. If the decision goes against Sizewell, it will make little sense to keep up the research. But a decision to build the Sizewell plant will leave unanswered the question how research and development in this field should be financed in relation to the numbers of nuclear power stations likely to be built in the years ahead. Walker or one of his successors is going to have to tackle that question soon. It should be more easily decided now that the UK Atomic Energy Authority has been made a trading fund, required to keep books as if it were a company.

The more immediate question of safety is similarly not absolute. But the fact that there is not yet an absolutely safe nuclear reactor does not imply that reactors cannot safely be accomodated with people in the same tract of land. Both Chernobyl and, earlier, the Three Mile Island reactor accident in the United States, were illustrations of how reasonably (but not perfectly) designed machines were induced to behave badly by their operators. The Three Mile Island accident is the more worrying because the operators acted sensibly in the period leading up to the accident. The moral, for those who would build reactors, is that much more attention even than at present must be lavished on the training, discipline and performance of the operators. Those in charge must be regarded as if they were airline pilots; accidents develop more slowly in nuclear reactors than in aircraft, but their consequences are greater. If Walker wishes to convince the British public that the Sizewell station could be safely built, he could do worse than to offer what countries such as Britain need in any case (and which is partly provided in the United States by the cumbersome procedures of the Nuclear Regulatory Commission), regular and public invigilation of the management of these plants.

But what should the minister and the British government decide? The Sizewell inquiry may have dealt with the need in Britain for a nuclear power programme as such, but the immediate proposal is that there should be a single pressurized-water reactor station five years or so from now. A single plant, especially one built in the present climate, is almost certain to be put together accurately and well. CEGB says it wants to build the machine; if it can confirm, in the light of information gathered since the Layfield inquiry ended, that the electricity is still worth having at the price, why not let it go ahead? The fact that the Labour Party has said that, if elected at the general election, it would "phase out" nuclear power, in any case argues for such a step-by-step approach. No doubt a decision by the government to sanction a single plant would be called provocative, but this is part of a legitimate disagreement; but committing public resources now to a long-term programme could risk an imprudent waste. What nuclear power in Britain and elsewhere now most needs is a show of cautious deliberation by sponsors willing to insist that its virtues are judged on strictly economic grounds and that its custodians, the plant operators and their bosses, shoulder their public responsibilities squarely. □

Shades of 1968?

The Chirac government in France seems to have made a hash of its intended university reform.

JUST why General de Gaulle's government took such fright when the students built barricades in the streets of Paris in 1968 has never been crystal clear. The immediate fear was that the survival of a nationalist government representing a deeply conservative people would itself be threatened if the students' anarchical protest at a modest package of educational reforms were allowed to run its course. But the whole of France was shocked by the violence of the students' anger at M. Edgar Faure's proposals for reforming the French universities. The police fought the students at the barricades, but the government withdrew its reform package, replacing it by a more anodyne reform.

It is odd that the new Chirac government should have fallen into the same trap. Not much seems to have changed in a decade and a half. Although the Chirac government's proposals for university reform have many objectives in common with those of 1968, the ostentatious orderliness with which last week's protest began shows that the students were not at first out to make a revolution. The death of one student at the weekend after a beating from the police has inevitably changed the mood. The government, having backed down half-way in stages, is likely now to have to give in altogether.

So are the reforms dead and buried? Not quite. The issues that have angered the students will surface once again in the next few years. So they should, for they are necessary reforms. The complaint against the Chirac government is not that it has made itself look weak and even foolish, nor that it launched the reforms in the first place, but that, with such an important objective to accomplish, it assumed that it would be sufficient to use its majority in the French parliament to get the reforms on the statute books, without attempting to carry academic opinion with it.

The truth, nevertheless, is that the French university system is one of the more enduring relics of the eighteenth-century revolution. The principles are grand, as has been their practice for much of the intervening period. Able teenagers at generally excellent secondary schools qualify for university entrance by means of the nationally uniform school-leaving examination and are then, in principle, free to choose the university at which they enrol. France being France, it is inevitable that the University of Paris should have had to grow fastest under the pressure of unregulated demand, but most good institutions have been overwhelmed by numbers. One of the government's now-postponed goals was that institutions should be more able to regulate external demand by setting their own entrance requirements; a by-product would have been an element of diversity within a system which, for all its virtues, remains too much of the same piece. Can that be bad?

So what went wrong? The French government's most obvious miscalculation is to have underestimated the passions attaching to the right to university education. But in the absence of positive proposals for helping students find and choose the institutions that would best meet their individual needs — financial assistance for students in France goes very little further than the nominal tuition fees (which were in any case to have been increased) — it is inevitable that the proposals would have seemed unjust to many potential students. That is why whichever minister of education brings these proposals back to the French parliament should take care to do so within a wider framework. France is no different from most other countries in Western Europe in facing a long period during which the student age-groups will decline numerically just at the time when greater skill and technical competence are more necessary than ever. The case for managing an element of diversity within the university system is strong, but not at the cost of the quality and even quantity of the students it trains. □

After the riots the politics

Chirac drops university reform bill to buy time

FRENCH Prime Minister Jacques Chirac appeared on national television on Monday morning to save his government. He announced the abandonment of the university reform bill that had led to demonstrations, then riots and finally the death of a student in Paris on Friday evening. Chirac's fragile coalition of right-wing parties was in danger of breaking up under the pressure of these events. Alain Devaquet, the universities and research minister, handed in his resignation on Friday night after the education minister René Monory announced, also on television, that he was to take over the steering of the bill; on Sunday, Alain Madelin, industry minister, said that the reforms "were not worth a life"; and others had begun to distance themselves from an administration that was rapidly being tainted as "anti-youth".

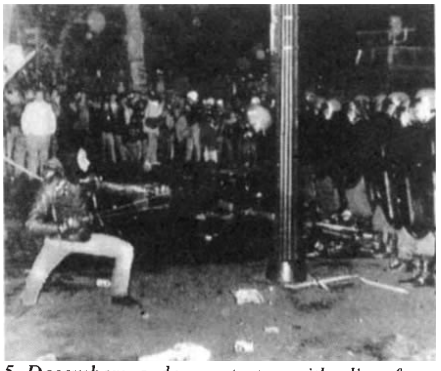
Just eight days earlier, Chirac had offered concessions to students on the three key points at issue — selection on university entry, local university degrees (they are at present national) and an increase in entry fees, which would vary from university to university But these proved not to be enough for student leaders who demanded complete renunciation of the whole bill.

Some, though, believe that the battle was won not by the students — who claimed to be apolitical in their protest — but by more violent elements that began to attach themselves to the previously peaceful demonstrations.

Where this debacle leaves the French universities and research is by no means clear. Devaquet had spoken of the "irony" of being tarred by students as an extremist, when his whole effort has been to tone down some of the more extreme libertarian elements of the bill that emerged from the right wing of Chirac's party and while he has successfully defended against that same wing other academic institutions such as the national research centre (CNRS). His resignation leaves a gap which Chirac will find difficult to fill.

But is Chirac's attempt at university reform dead?. Chirac is an authoritarian (if realistic) prime minister, who said only a few days ago that he had been elected to "govern", so a new bill may yet emerge once tempers are cooled.

As to why the French students — and high-school children — were so determined to resist selection, and increasing university autonomy, the answers are more social and psychological than political.



5 December: a demonstrator with sling faces riot police.

Selection does operate in French universities already, but it bites late on entry to the second and third years. It is made by university staff on new grounds; so a schoolboy or girl who has failed to make good progress at the very authoritarian French schools has a second chance in the first year of university (provided he or she passes the school-leaving *baccalaureat*). Pre-university selection means denying that second chance; and that can mean the dole queue.

The previous government planned to increase the entry to the first year of university, and making it explicitly a year of 'orientation' that could lead to jobs or other educational institutions as well as onwards in the university; but it failed to provide the resources to accomplish the task and created resentment among university staff. This government attempted to please that staff. But it forgot about the students.

Robert Walgate

Vera Rich was in Paris when the students' patience ran out . . .

PARTICIPANTS in last week's clashes between French university students and the police were at a loss to explain how matters had got so out of hand. What was planned as a peaceful, silent, non-political protest on Thursday escalated into disturbances which, on Friday night, culminated in the death of the 22-year-old student Malik Oussekine. Certainly the Thursday demonstration was "politicized" by participants who, contrary to the organizers' appeals, brought political banners. But a high police profile at the march aggravated the situation. By Friday, the nationwide protests had taken on a frankly political tone, and on Saturday trades union leaders were being asked to sit in on the meetings of student delegates from all over France, and calls were going out for a one-day general strike.

The government's reactions — as so often in such situations — were too little, too late. The first, partial climb-down over the bill might have been effective on Thursday, before violence developed but it came only on Friday, a few hours before Oussekine's death. Security Minister Robert Pandraud's statement, on Saturday, that if there had been "individual faults" on the part of the police, sanctions would be imposed, was coupled with a refusal to dissolve the special police 'anti-vandal' motorcycle squadds introduced early on Saturday morning.

Interestingly, by Saturday night, even students involved in the growing disorders were convinced that the genuine protests had been "manipulated", although opinions differed as to whether the alleged manipulators came from the extreme right or the extreme left.

From the beginning, there has been an emotional and in some degree irrational element in the students' and *lycéens'* opposition to what, viewed more logically, might have been seen as an unpleasant but necessary reform, given the current economic state of France and the problems of graduate unemployment. But *lycée* scholars have grown up considering it an inalienable right that all school leavers who pass the *baccalaureat* can automatically go to university, while restrictions on choice and combinations of subjects, and the possible development of a hierarchy of excellence among the universities, seemed a direct affront to liberty, equality and fraternity. Some students, too, seem to have been stung into their decision to take part in the demonstrations by the accusation, voiced on French television only a couple of weeks before by Daniel Cohn-Bendit, the 1968 student leader, that the present generation of French students are "apolitical".

In reality, as several university lecturers have been at pains to point out, the hierarchical principle has long been established *de facto* in France, with prospective employers at least as interested in where an applicant studied as in his or her subject and academic record. The new law, they said, would simply validate the existing system. But as early as Friday afternoon, and particularly since the death of Oussekine, these same lecturers began privately to adopt a different stance. "I do not know whether I am on the side of the students over the law", one said, "but I am certainly not on the side of the police against the students!"

Vera Rich

European summit

Unanimity but little of substance

EUROPEAN leaders confronted with the massive overproduction and costs of the Common Agricultural Policy and the need for technological cooperation to face competition from the United States and Japan, laboured mightily to produce a mouse last week: an anodyne communiqué, the result of the London summit marking the last month of Britain's six-month presidency of the European Communities, which failed to mention agriculture at all, put off budget questions to a later summit, and included lukewarm support for research in Europe only on the last-minute insistence of the French President, François Mitterand.

This time, the presidents, prime ministers and chancellors of the 12 states of the European Communities were in no mood to tackle controversial topics, at least in public. Foreign Office sources in London indicated not entirely in jest that the present disablement of the US presidency over the Iran arms sales affair, and the confusions over US defence policy, obliged Europe to present a united face to Moscow: squabbling in the Community would not help. Moreover, Chancellor Helmut Kohl of West Germany faces general elections next month, so this was not the best time to ask him to be hard on his farming voters. Prime Minister Jacques Chirac of France was preoccupied with the student protests (and late on Friday, the death of a student) in Paris; and the United Kingdom, meanwhile, did not want to bow out of the presidency on a sour note after its efforts these six months to prove itself a hard-working, constructive member of Europe.

The resulting bland unity was not good news for the research programme of the Communities' think-tank and administration, the European Commission, which the Commission has proposed to double. Jacques Delors, Commission president, had hoped the summit could send a strong signal in favour of the programme to the research ministers, who were to meet on Tuesday. During the summit, Delors emerged to say that many delegations had stressed the importance of the Commission's 'framework programme' for research and "approved it totally". Finally, however, the summit communiqué merely "urged" research ministers to agree, but gave no indication on the level of spending on which agreement should be reached, the matter which seemed likely to be the final sticking point on Tuesday between the big three — Britain, France and Germany — which oppose the proposed increases, and the rest, which support them.

President Mitterrand, pressed on the research question after the summit — as he had been the first to introduce such matters to summitry — claimed credit that any clause on research at all had been introduced into the communiqué. The British Prime Minister, Margaret Thatcher, also questioned on research, avoided the matter of budgets and stressed the importance of agreement on standards for high technology equipment. She singled out the need for a common standard on digital cellular radio. This relatively small item was also the only technology explicitly mentioned in the communiqué for action, leaving the question of whether the leaders might actually have had in mind the need for common standards in the more ramifying systems of integrated digital services networks.

The European leaders also reached agreement to designate 1989 European Cancer Information Year to develop a "sustained and concerted information campaign . . . on the prevention, early warning and treatment of cancer". No mention was made of possible concerted action on tobacco advertising, but the ministers undertook to "follow up" the recommendations of a committee of European experts which is due shortly to report on cancer prevention to the Commission. A programme on AIDS (acquired immune deficiency syndrome) was also agreed, but this was only to compare notes on actions taken by the different member states in public information and prevention. The possibility of "further cooperation in research" on AIDS was to be considered by health ministers, the leaders agreed. **Robert Walgate**

Independence for UK space centre?

THE British National Space Centre (BNSC) is pressing the government to give it financial independence and to allocate an annual space budget of about £200 million without channelling it through other government bodies. The proposals for the budget and plan are at present being discussed at cabinet level.

The proposal for direct support accompanies the 15-year BNSC plan, and would allow more flexibility in managing the UK space effort. At present, the centre is funded by the Department of Trade and Industry, the Department of Education and Science and the Ministry of Defence.

The BNSC plan would increase spending on space, now £108 million, to about £200 million a year within three or four years. The plan has four principal ingredients and puts heavy emphasis on European collaboration. The first part of the space plan recommends strong British participation in the US Space Station and in a polar platform.

The second plank of the plan will ensure the development of new launchers, independently of the United States. Ariane V would form part of that project together with cheaper forms of space transport, including the European space shuttle HERMES and the British Aerospace/Rolls Royce space-plane HOTOL. The plan also includes an expanded space applications programme — Earth observation, microgravity and the like — as well as more space science research. Studies of solar plasma, geophysics and astronomy would also feature in this programme. Bill Johnstone

Super-collider still fighting for funds

Washington

LAST week the threatened Superconducting Super-Collider (SSC) project cleared one of its major political hurdles by gaining approval from John Herrington, the first Energy Secretary to endorse SSC since the agency began work on the programme five years ago.

It is a change of heart as only a month ago Herrington claimed that his mind was not made up on the advisability of the project. But because of the enormous cost, now estimated to be about $3,000 million, formal approval would have to come from the White House Office of Management and Budget (OMB). According to sources at the Energy Department, Herrington has not yet formally met top OMB officials.

Alvin Trivelpiece, director of the Energy Department's Office of Energy Research, has been trying to get other countries to participate in SSC, but such commitments are difficult if the United States is hesitant. Trivelpiece says Japan may invest $500 million in the project. The Soviet Union is another potential participant. For the time being, potential partners from Western Europe have their hands full with LEP and DESY.

While approval from the Reagan administration is a necessary step for the project to proceed, it is far from sufficient. Many believe that the brake on spending achieved by the current Congress was achieved by adopting financial tactics such as delaying payments and selling assets. Any further reductions will require real programme cuts.

Congressional enthusiasm for the National Institutes of Health shows no sign of diminishing, and the administration is rumoured to be considering a large increase in the budget for the National Science Foundation and a request for more funds for the Strategic Defense Initiative. It is uncertain whether spending on SSC will prove to be any more politically attractive. Joseph Palca

Amazon rainforests

Brazil invites RGS explorers

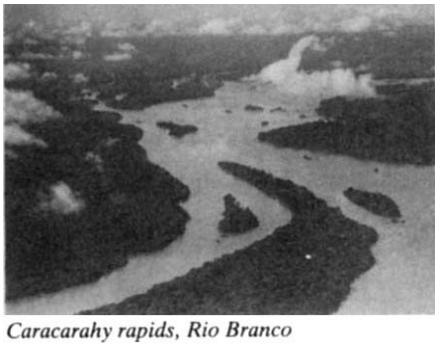
THIS week the first of 60 scientists depart for Brazil to make the final preparations for the biggest European-led research expedition ever to study an Amazon rainforest. The project team, led by Dr John Hemming, director and secretary of the Royal Geographical Society (RGS), will be based on the island of Maraca close to the watershed between the Amazon and Orinoco basins.

The invitation was made to RGS two years ago by the Brazilian Environmental Secretariat (SEMA), but it has taken since then to organize and raise from a mixture of industrial, private and government sponsors the £125,000 that will fund the basic needs of the year-long project.

Hemming says "we jumped at the invitation since, although Brazil contains almost two thirds of the world's surviving rainforests, it is difficult for foreign scientists to obtain permission to work there."

Half of the teams will come from the United Kingdom with Brazil making up most of the remainder. Maraca, an uninhabited island 60 km by 25 km, is covered in forest, contains a lake and has swamps and seasonally flooded areas. As a result of the expedition, SEMA, whose permanent research station on Maraca is at the disposal of the RGS team, will acquire a complete survey of the island, including an inventory of the animals, insects and plants, to help plan future research.

RGS says its five research programmes



Caracarahy rapids, Rio Branco

have "been welcomed by the Brazilian authorities because of the practical benefit to both farmers and conservationists". They involve an ecological survey, the development of management plans for the reserve, an examination of forest regeneration and the behaviour of soil moisture, a medical-entomological review and a survey of land development.

The Royal Botanic Gardens at Kew in London will oversee the ecological survey which will record, classify and collect hundreds of species of trees, plants, vines and ferns with the help of specialists from the Royal Botanic Garden in Edinburgh and the New York Botanical Garden.

There will be an inventory of hundreds of species of fauna, particularly snakes, frogs, bats, monkeys, insects and birds. Studies of forest regeneration could help restore destroyed areas to their original state. According to RGS, thousands of hectares of rainforest are felled every year for farming, settling, cattle ranching and mining. But a wasteland usually results, claims RGS, and much of it is soon abandoned. Farming proves impossible because of poor soils and intense parasite and weed activity. Once damaged, the rainforests do not regenerate and the research will try to find out why.

Another research programme will investigate how the litter (or debris) on the forest floor is derived from the foliage and decomposes, the dynamics of the rainfall as it drops from the roof or canopy of the forest and the behaviour of that moisture with the soil. Research on endemic tropical diseases will focus on onchocerciasis and leishmaniasis.

The land development survey will study the settlements within reach of the research camp. According to RGS, millions of settlers are moving to the frontiers of the Amazonian forests in a desperate search for land. "This occupation and disturbance often occurs too fast for there to be any planning based on scientific information. As a result, many pioneer colonists fail."

Bill Johnstone

Space shuttle

Money-saving move on fuel tanks

Washington

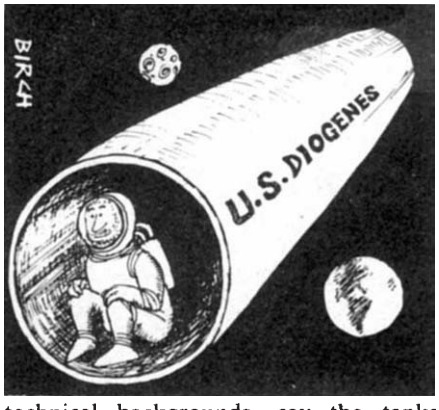
THE idea of making some use in orbit of each space shuttle's huge external fuel tank is not new, but prospects for its realization may have come closer. Hitherto, the fuel tanks have been allowed to burn up in the atmosphere. But now a new company, External Tanks Corporation, wants to make them into warehouses. The company is seeking private capital to finance what it calls the first private "warehouse in space". The project has the backing of the University Corporation for Atmospheric Research (UCAR), a consortium of 57 major research universities. According to Thomas F. Rogers, chairman of External Tanks Corporation and a former member of NASA's space programme advisory council, the first habitable tank could be in orbit by 1993.

The shuttle's external tank carries liquid hydrogen fuel and oxygen for the main engines, and is usually jettisoned just before the shuttle reaches orbit. Its diameter is more than the width of a Boeing 747, and its length is equivalent to a 13-storey building. Twenty-four have so far been discarded, each costing $30 million. To take the tank into orbit would require only an extra 2,000 lb of fuel, according to Randolph Ware, president of the new company.

The company, which is 80 per cent owned by the UCAR Foundation, a non-profit organization established by UCAR, is seeking capital from "companies that have a long-term strategic interest in the exploitation of the space environment", including hotels. The National Aeronautics and Space Administration (NASA) is apparently taking the company's plan seriously. Several studies have recommended that the idea be investigated, and NASA has now formed a special committee to look at it, chaired by general manager Philip Culbertson.

Culbertson says the "very interesting" idea is "feasible", but adds that NASA has not yet looked into all the technical details and that it would in any event leave the business side to External Tanks Corporation. NASA has some plans of its own for using the tanks: it recently awarded a contract to Martin Marietta Corporation to study turning an external tank into a gamma-ray imaging telescope.

Randolph Ware, president of External Tanks, and Rogers, both of whom have



technical backgrounds, say the tanks would need minimal modifications on the ground. Once in orbit, the tank would be detached from the shuttle and fitted with a small thruster that would keep it in stable orbit. Subsequently it would be purged, pressurized and fitted as as "the equivalent of high-cost commercial space", according to Ware, who seems to have no doubts that commercial companies will want to rent.

Among the plans being discussed is one for a fully pressurized permanently inhabited laboratory/factory, Labitat, made from two tanks connected together. The operating cost of Labitat is calculated to be so low that accommodation would cost as little as $500 per person per day by the 1990s, roughly twice the cost of first-class hotel accommodation in the United States.

Tim Beardsley

Climatic prediction

Could this be an El Niño?

Washington

SOMETHING is happening in the equatorial Pacific Ocean from the dateline to the west coast of South America. Some believe that the climatic upheaval known as El Niño is in progress, others believe one is ending, some call the events an ENSO episode (El Niño/Southern Oscillation); others claim that all that is happening is that a travelling wave of water is moving through the Pacific from west to east. But there is agreement on one point: something is happening.

The events of interest relate to a warming of the sea-surface temperatures in a band from the dateline eastward, a rise in the water temperature measured from piers along the Peruvian coast, and slackening of easterly sea-surface winds. Whether these events should be termed an El Niño depends on definition. Originally, says George Philander of the National Oceanic and Atmospheric Administration (NOAA), an El Niño referred to warm current flows along the coasts of Ecuador and Peru in January, February and March, and the resulting impact on local weather. But James O'Brien of Florida State University in Tallahassee says no two events have been the same, making the 'classic' definition of little value. The second name, ENSO, more generally refers to events from the mid-Pacific to the South American coast, taking into account the irregular oscillation in pressure between the east and west Pacific.

The National Meteorological Center's Climate Analysis Center (CAC) issued an advisory note on 10 November stating that the warming of sea-surface temperatures in the equatorial Pacific "resembles the warmings associated with ENSO episodes". But the centre was unclear whether El Niño conditions would result. CAC's Don Rodenhouse says the next circular will probably appear next week based on November data, and is likely to confirm that Pacific warming is spreading.

The National Weather Service, CAC's parent agency, is cautious about using the term El Niño. Last March it announced that conditions that could lead to an El Niño were developing. Eugene Rasmusson, now retired from CAC but responsible for the March bulletin, says the warming trend failed to materialize. Many people, mindful of the economic disasters blamed on erratic weather from the 1982–83 episode, mistook the possibility of the event for the event itself, provoking some hyserical behaviour. Now the Weather Service will move very cautiously before making any firm statements about El Niño.

But the government's caution does not extend to the academic community. There is "absolutely no question in my mind that there is a modest El Niño under way", says O'Brien, who is chairman of the ENSO alert group that advises CAC on its Pacific forecasting.

Several members of the group had developed models as early as January predicting the ocean temperatures now observed. But O'Brien says none of them felt that this year's episode would match the 1982–83 episode in scale. From his own model, O'Brien predicts that sea-surface temperatures along the Peruvian coast may soon begin to decline.

Francisco Chavez of Duke University says his data show that decline is taking place based on pier temperatures from Paita, 5° south of the Equator. But Chavez data also show that pier temperatures for the start of November were well above normal.

Tim Barnett of the Scripps Institution of Oceanography has made a statistical model that predicts the behaviour of sea-surface temperature in the equatorial Pacific. Barnett's observations of temperatures in the region show about 1.2 °C above normal. His model predicts that the current warming trend will continue for another month or more.

Mark Cane, Stephen Zebiak and Sean Dalan of Lamont–Doherty Observatory in Palisades, New York, developed a deterministic model based on a coupled ocean–atmosphere system (see *Nature* **321,** 827; 1986). Cane says that if the model is correct, the current warming will have already peaked. His analysis of the latest data from the region supports that view, although the temperature could rise again. Barnett says the exciting thing will be to see how well the models behave in predicting events and to understand why they work as well as they do.

Predicting an El Niño is not necessarily of value in itself. "The term El Niño is useful like the term winter is useful", says NOAA's Philander. The significance of El Niño episodes for global weather is still not well understood. Moreover, Philander is not convinced that the rise in coastal temperature is due to an ENSO episode — a travelling wave, known as a Julian–Madden wave, may be bringing the warm water from the west, rather than a perturbation of the southern oscillation. One measure of the southern oscillation, the pressure difference between the island of Tahiti and the city of Darwin in Australia, has failed to show the difference normally associated with ENSO episodes.

But Rasmusson believes a weak-to-moderate ENSO episode is indeed under way, and the Tahiti-minus-Darwin pressure index is just one of the factors still not clearly understood.

Joseph Palca

Biotechnology patents

Don't say just what you mean

Rehovot

AN extremely inventive patent covering biologically-engineered "human pink-spider toxinase" was recently presented for examination by Wiezmann Institute's Professor Hermona Soreq and Professor Avigdor Shaefferman of Israel's Biological Research Institute. But the toxinase and the cultured cells yielding a miracle cure for the dread spider bite were all the figments of the inventors' imagination, designed to illustrate the difficulty of patenting intellectual property.

The pink-spider toxinase patent, together with an equally inventive suit claiming patent infringement, was part of an unusual experiment carried out at an international conference on legal aspects of biotechnology held in Jerusalem. The invention was presented for examination by Mr Joel Tzur, the Israel Commissioner of Patents, and it was finally 'approved' after a mock appeal heard by the deputy president of the Tel Aviv District Court, Judge Ellyahu Winograd. Among those attending the Jerusalem meeting were scientists, jurists, patent attorneys, representatives of patent offices and industrialists. The so-called biolegal meeting was the brainchild of Tel Aviv District Court Judge Shoshana Berman. The importance of using precise legal terminology in patent applications became clear at the conference. For example, delegates concluded that it is not appropriate to describe a potentially patentable protein fragment as being of a specific length; instead, it should be described as being *essentially* of a specific length. That extra word makes it far less likely that the application can be successfully challenged. "I learned", Professor Soreq concludes, "that the exactitude demanded of scientists may be counterproductive when they are applying for a patent."

The Israel Patent Office has been issuing patents on genetically engineered microorganisms even before they were patentable in the United States. But Professor Haim Aviv claimed that Israeli authorities have been too liberal in granting broad patent protection to foreign biotechnology companies, which, he charges, are less interested in manufacturing their products than in stifling Israeli innovations.

Aviv, himself deeply involved in the business side of biotechnology, hopes that Israel will soon adopt the US principle of a grace period, so that local researchers, like their US counterparts, will have a full year to patent their findings after publication in a scientific journal.

Nechemia Meyers

Bhopal

Liability for disaster still disputed

New Delhi

THE question of compensation for victims of the Bhopal disaster still remains unresolved on the second anniversary of the leak of methyl isocyanate (MIC). The legal wrangle has taken a new turn with the Union Carbide Corporation (UCC) alleging that it did not own the plant and therefore is not liable for the incident on 2 December 1984.

In a 179-page document filed in the court of Bhopal's district judge, UCC said the Indian claim was based on the "erroneous assumption that the plant belonged to, and was designed operated and maintained" by UCC. It said the plant was designed and constructed by Union Carbide of India Limited (UCIL) in consultation with Indian engineering companies, and that UCC provided only the process design of certain parts of the plant in accordance with the provisions of the design transfer agreement approved by the Indian government. The agreement, according to UCC, specifically stated that UCC "shall not in any way be liable for any loss, damage, personal injury or death resulting from or arising out of the use by UCIL of the design packages". UCC said it was only a shareholder of UCIL which alone "is exclusive owner of the Bhopal plant".

Apart from seeking to transfer responsibility for the plant to UCIL, the US company has denied that the emission of MIC from the plant occurred as a result of defective plant conditions. It said that the incident occurred because large quantities of water were deliberately introduced into the MIC tank by connecting a water hose to the opening from which the tank's local pressure indicator had been removed.

According to UCC, the pressure indicator was in place on 30 November (two days before the accident), but on the morning of the accident it was found to be missing and a water hose was noticed lying on the top of the tank. UCC claims that the plant supervisor's daily notes included a sketch showing that water had entered through a connection to the pressure indicator. It said that somebody deliberately intending

to introduce water into the tank could, within a few minutes, close the isolation valve, remove the pressure indicator, connect a water hose from a water service drop only 10 ft away and open the three valves necessary to introduce water.

UCC said that the facts "establish that a water connection to the tank was found and concealed by the operators and supervisors on the third shift" and that the Central Bureau of Investigation and the scientific advisers, with all the evidence available, "apparently desired not to reach the conclusion that the event was caused by a deliberate act". This "had the effect of aiding the persons involved in maintaining their deception", according to UCC.

The government of India maintain that UCC was liable for the consequences of the incident and is suing the corporation for $3,000 million in damages, saying that 2,347 people died and 30,000–40,000 were seriously injured in the disaster.

The legal arguments are expected to be prolonged, offering no immediate help for the survivors, many of whom, according to the Indian Council of Medical Research, continue to suffer from chest ailments and are without jobs. The Madhya Pradash government has so far spent more than $40 million on health care and rehabilitation; UCC and UCIL have made available about $16 million.

K.S. Jayaraman

Space Station cooperation

Money the root of conflict

Washington

DISAGREEMENT continues over arrangements for European participation in the planned US Space Station. European officials are pressing the United States for assurances on operating procedures and cost-sharing, asking that these be specified in a formal intergovernmental agreement. The United States, however, wants many of the details left to a memorandum of understanding between the National Aeronautics and Space Administration (NASA) and the European Space Agency (ESA). At issue are the financial returns the partners can expect from the collaborative venture and the extent of European autonomy in planning.

Several high-level meetings have been held recently between government and space agency representatives from different countries, including Japan. The negotiations are taking place against a background of uncertainty about the cost of the whole project. Recent NASA estimates suggest that the original figure of $8,000 million will fall far short of requirements unless the station capabilities are reduced. All sides say they want to reach final agreement as soon as possible, frustration is evident. One European official commented that Europe knows from experience that the United States does not take memoranda of understanding to be binding, a reference to the US decision some years ago to scale down its participation in the International Solar Polar Mission.

Officials also concede that serious differences remain over key aspects of management and sharing of costs and space station resources. Europe remains committed to its own $2,000-million pressurized module called Columbus, but is unhappy with the US view that that should limit its share of the total returns to 15 per cent. NASA is required by congressional directive to recoup 80 per cent of space station benefits; one important question, according to Richard Barnes, director of NASA's international affairs division, is whether Columbus should be considered an integral element of the space station system or an adjunct. The same applies to the European plan to develop a man-tended free flier, on which it has insisted despite US misgivings. Questions remain about how shuttle or Ariane launches and provision of payload specialists should appear in the accounts.

Europeans are also unhappy with a US congressional requirement that the United States should be responsible for materials science operations in its modules, leaving biological sciences for the Europeans. Officials say, however, that despite the congressional directive, there is now understanding that Europe will participate somehow in materials science research on the space station and be allowed some of the benefits. It is agreed in principle that commercial companies will have laboratory units that will sometimes be closed to general access in order to keep proprietary information confidential.

Both sides say they want final agreement on both the memorandum of understanding and the intergovernmental agreement by the end of the year or by very early in 1987. The European Council of Ministers is meeting to discuss space next spring, and would like a definite framework established by then; NASA wants to release a final request for proposals for the four main station work packages in January so that detailed development can begin in earnest next summer.

Tim Beardsley

Monoclonal patent

Hybritech versus Abbott

Washington

HYBRITECH Inc., a subsidiary of Eli Lilly, is suing yet another company for using monoclonal antibody assays in its diagnostic kits. Hybritech has now decided to sue Abbott Laboratories on the grounds that some of Abbott's diagnostic tests may violate its patent for 'sandwich assays'.

The patented Hybritech process uses two monoclonals, each specific to a different site on the same antigen, to make an antibody–antigen–antibody 'sandwich'. This detects small quantities of antigen, allowing early diagnosis of disease.

The interest of the new suit is that Hybritech is asking not only for damages and a permanent injunction against Abbott, but for a preliminary injunction that would prevent Abbott from selling its materials even before the case has come to trial. To persuade a judge, Hybritech will have to prove that continued violation of the patent would cause it "irreparable damage". Patent attorney Jorge Goldstein, of Seidman, Sterne, Kessler and Goldstein, says the chance of winning such an injunction, previously not high, is greater in the present pro-patent legal climate.

The disputed products include Abbott's pregnancy tests and some of its hepatitis and cancer tests, which are sold only to physicians or hospitals and comprise 7 per cent of Abbott's sales of diagnostics. Chuck Weber, a spokesman for Abbott, says that their test for the AIDS (acquired immune deficiency syndrome) virus, which uses polyclonal, not monoclonal, antibodies, is not affected by the suit.

If Hybritech wins this case it could also go after other companies using the sandwich assay technology. Syncor, in Sylmar, California, and Ventrex, of Portland, Maine, both have tests for human chorionic gonadotropin that use Hybritech's sandwich assay process to test for pregnancy. And Centocor, of Malvern, Pennsylvania, has just received US Food and Drug Administration approval for its CA 125 ovarian cancer blood test using the technique.

Hybritech's new suit may have arisen from the latest turn in its earlier suit against Monoclonal Antibodies, Inc. In 1984, Hybritech's suit failed and the patent was ruled invalid due to obviousness, but that decision was overturned, on Hybritech's appeal, last September. This decision returns the case to the federal district court in San Francisco, to determine if Monoclonal Antibodies' products infringe the patent for sandwich assays.

But David Brezner, chief counsel for Monoclonal Antibodies, intends to petition the Supreme Court over the decision upholding Hybritech's patent. There is some possibility that the justices may hear the case, as similar suits will certainly arise over other biotechnology patents in future. Brezner said the Abbott case "will not have a major impact" on his company's litigation with Hybritech.

However the Hybritech cases are settled, their handling by the courts will influence future biotechnology patent cases. Jorge Goldstein predicts a high volume of cases for the coming years. But ultimately, he believes, when several patents have been upheld, fewer companies will be tempted to infringe existing patents.

Carol Ezzell

AIDS research queried

Skulduggery at the lab bench

Washington

"DISTRUST and dislike" between AIDS (acquired immune deficiency syndrome) researchers is the probable explanation for incidents of tampering at the federal Centers for Disease Control (CDC) AIDS programme laboratory in Atlanta, Georgia, according to a special independent assessment by the Institute of Medicine. The assessment concludes that a "lack of strong scientific leadership" at the laboratory has led to dissatisfaction and low morale among researchers.

The assessment, requested after allegations in the press of data suppression and tampering, says the committee could not rule out minor tampering with equipment. Reports received "suggest that some unusual incidents did in fact occur".

The committee, chaired by Dr Julius Krevans of the University of California at San Francisco, found the AIDS laboratory only "moderately productive" and recommended that CDC make institutional changes to separate basic and applied research. It also says that the director of CDC, Dr James Mason, should "take a substantial and visible role in the implementation of the new arrangements", and called for a review of authorship policy for AIDS research, a "locus of discord". Some individuals said their contributions were not recognized, and others felt that individuals not involved were credited. Some changes have already been made.

Another allegation made in the *Atlanta Journal* and *Atlanta Constitution* that research data were suppressed was not confirmed. The data in question were from experiments investigating the efficacy *in vitro* of Nonoxynol-9, a spermicide, in killing the AIDS virus. Early drafts of a manuscript published in *The Lancet* in December 1985 (Hicks *et al.* **ii**, 1422–23) had been submitted for internal review at CDC in January 1985, but the final version was not submitted until November.

Krevans' committee concludes, however, that although it is understandable that critics wanted to avoid unwarranted extrapolation of the results, "it is not clear why the revision required six months and clearance seven weeks".

Tim Beardsley

New hepatitis-B vaccine launched

Brussels

COMPETITION in the potentially lucrative market for vaccines against hepatitis-B virus is hotting up. Following close on the plans of a Korean company to produce a dollar-a-shot plasma-derived vaccine (*Nature* 324, 399, 1986), SmithKline Biologicals last week received official approval (in Belgium) for its new recombinant DNA vaccine. Already extensively tested on volunteers, the SmithKline vaccine will compete directly with a similar recombinant DNA vaccine produced by Merck.

The recombinant vaccines are composed of the coat protein ('surface antigen') of the hepatitis-B virus, made in and purified from yeast. The plasma-derived vaccines are made from particles containing surface antigen that are present in the blood of chronic carriers of the virus. As far as SmithKline has tested it (15,000 doses on 6,000 volunteers), the new vaccine appears to be equally effective and can provide a boost to people immunized with the plasma vaccine. Immunity takes about 4 doses and probably lasts about ten years.

Another antigen that may be important in hepatitis-B virus infection is known as Pre-S; some reports say that antibodies to this antigen prevent the virus attaching to the target cell, others that they predispose a patient to become a carrier. Because of this uncertainty, Pre-S was not included in the current vaccine, but according to Dr Alain Goudeau, who has been conducting clinical trials for the company, they have plans to make and test a vaccine that does contain it.

A course of vaccination with the plasma vaccine at present costs around £90 and SmithKline intends to keep the price levels about the same for the time being; the vaccine has been on sale in Singapore for some months at about 20 per cent less than the plasma-derived vaccine. As the sales increase, it is expected that the price will drop; the company has already invested in a new 1.5 m³ fermenter to increase volume. SmithKline is confident that it can compete with Merck on price and availability, but the competition with the Korean vaccine on price will probably be the deciding factor.

Rebecca Ward

Japanese technology

Old computers may prove best

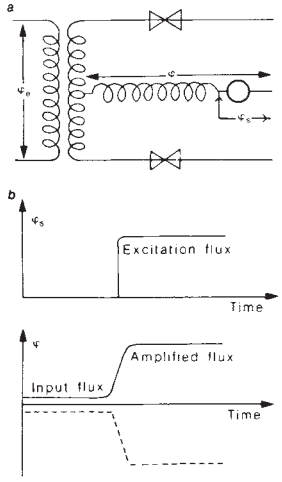
Tokyo

JUST at the time IBM (International Business Machines) drastically scaled down its research and development of Josephson junctions, Professor Eichi Goto of Tokyo University plunged into investigation of the supercooled logic devices, an idea originally developed by him more than 30 years ago. Now, three years later, Goto has been awarded about Y15,000 million ($92 million) by Japan's Research and Development Corporation under the ERATO programme to develop a Josephson junction computer that will process information in the form of minute magnetic fluxes.

The 'quantum magnetic flux logic' project will span five years and will be carried out by three research groups: Hitachi Central Research Laboratory will build the hardware, a team at Mitsui Shipbuilding will develop the software and a group at ULVAC in Chibasaki near Tokyo will be responsible for the magnetic shielding and refrigeration.

Goto achieved worldwide fame in the 1950s with his patenting of the parametron, a resonant electrical circuit that can store binary information by utilizing the two possible phases of oscillation of the resonating current. But although several Japanese companies, including NEC and Hitachi, developed parametron computers, these machines were soon superseded by much faster transistorized computers in the 1960s, and parametron computers can now be found only in science museums.

Nevertheless, Goto believes that a Josephson junction computer using a new type of superconducting parametron could operate at 10 GHz, a hundred times faster than present-day computers, while consuming a thousandth of the power of other Josephson junction devices. The secret lies in using magnetic flux, rather than electrical current, as the conveyor of digital information.



A d.c. flux parametron.

The d.c. flux parametron outputs an amplified magnetic flux the polarity of which is determined by that of a small input signal (see figure). The two possible polarities, positive or negative, can represent the 0s and 1s of binary logic, and by linking parametrons together magnetically, all the basic logic circuits of a central processor can be built. And in fact Hitachi has already linked together three such parametrons to make a clocking device that operates at about 2 GHz.

One major problem, however, remains. The magnetic fluxes on which this computer is supposed to operate are so tiny, on the order of nanogauss, that the computer must be completely shielded from external magnetic fields, such as that of the Earth and underground electric trains.

Undeterred, Goto has already set about trying to build the "perfect magnetic shield". In theory, a superconducting shield encasing the computer could shut out the external magnetic field. But in practice magnetic flux is pinned to the superconducter during cooling, setting up localized internal fields. Goto's plan is to 'sweep' the shield clean of pinned flux with a laser beam, but so far without success.

During preliminary experiments with a superconducting cylinder made of lead, more than 99.9 per cent of the laser's light was reflected by the shiny lead surface and a 'hot spot' for collecting flux failed to develop. Maintaining the cylinder black did not help much either — the paint peeled off on cooling.

If successful, Goto will be able not only to protect the computer from magnetic fields but will also have the ideal detector for magnetic monopoles, which were predicted by the late P.A.M. Dirac.

David Swinbanks

Indian resources

National water policy at last

Bangalore

INDIA has been trying for many years without success to evolve a comprehensive national water policy, but this goal is now a step nearer.

The chief need is for efficient and integrated management of water resources and more effective use of irrigation facilities. Mr B. Shankarananda, Minister for Water Resource Management, says the new policy, which is now getting its final touches, is intended to make available safe drinking water for more than half a million villages and also to link the Himalayan and peninsular rivers so that surplus water can be transferred to areas that need it. Cost-effective techniques for finding water in drought-prone areas, particularly in the north-eastern part of the country, are also being sought.

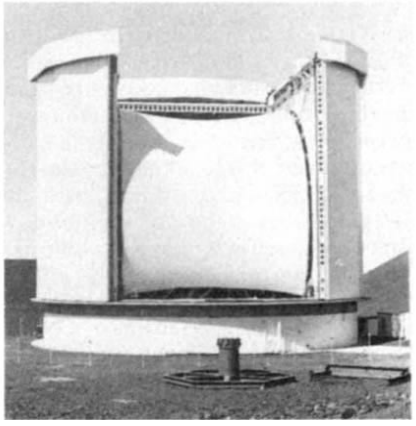
Under the new policy, both central and state governments would share the cost of creating water channels, dams and drinking-water sources. Water would be declared a national asset so as to end interstate disputes.

In the meantime, the Indian Planning Commission has announced that irrigation and flood control projects that do not include plans for 'ancillary aspects' such as soil conservation and protection of the ecosystem will not be permitted.

Water-borne diseases account for the loss of 73 million man-days in India each year. In 1980, it was estimated that out of 570,000 Indian villages, more than a quarter have no source of drinking water within 16 km. Although Rs77 million had been recommended by a working group for rural water supply and sanitation in 1986–90, only Rs34.5 million was allocated.

Radhakrishna Rao

UK–Dutch telescope



LAST week the James Clerk Maxwell Telescope, in the final stages of its initial commissioning tests, received its first radio signals from space. The £20-million instrument, installed at Mauna Kea Observatory on Hawaii, is the largest telescope in the world operating in the submillimetre region of the spectrum.

The telescope, jointly funded by Britain's Science and Engineering Research Council (SERC) and the Dutch Nederlandse Organisatie voor Zuiver-Wetenschappelijk Onderzoek (ZWO) will become fully operational next year.

The first radio signals to reach the telescope last Thursday were those from the Moon, Jupiter and Mars. The telescope has been specifically designed, however, to penetrate the dense regions of interstellar gas, which are opaque to optical radiation. At the infrared and millimetre wavelengths, these regions are transparent.

Bill Johnstone

Italian research reorganization

SIR—Your article on research in Italy (*Nature*, **323**, 751; 1986) depicts the Italian government as working quickly at reorganizing the national research institutions.

As a researcher with 18 years' experience and an employee of the National Research Council (CNR) which, with its 6,189 employees, is the largest of the government research institutions, I would like to express my incredulity at the effects that this difficult reorganization will produce. By the time the reform of CNR is approved by parliament, the most talented and best qualified researchers will have fled to more remunerative positions in universities and industry.

It may be true, as reported in the article, that the average cost per researcher is about 130 million lire. (Please note that public research institutions in Italy employ a mere 14,000 of the 112,884 cited.) But the annual salary of a senior CNR researcher — whose competence, for the information of non-Italian readers, is at the level of a full professor at a university — is no less than 13 million lire, about £6,500.

It is with a sense of shyness mixed with a sprinkling of rage that we watch the CNR president, Professor Luigi Rossi-Bernardi, and the Minister of Research, Senator Luigi Granelli (whose joint efforts to save CNR must be acknowledged), praise the abundant productivity of CNR scientists and the consequent healthy state of CNR. Most of us have spent several years abroad, often working in renowned international research centres, and are all aware of the conditions in which we continue to do research in Italy, and of how they compare with conditions in Britain, the United States and Japan.

Interestingly, the fate of CNR researchers and their hopes of better salaries now seem to be linked to the approval of a government bill (not the reform of CNR) originally devised to exempt the wages of workers at the Mint and of air traffic controllers from the restraints of the anti-inflation finance law, which limits the increase of public employees' wages to 6 per cent a year.

It is well known that Italy suffers from frequent government reshuffles. Experienced observers think it likely that there will be another next March. Whereas I am sure that most of the many "million million" lire projects cited in your article will be funded, I doubt whether the contract for CNR employees (which expired on 31 December, 1984) will be renewed by the end of this year.

MARCELLI G. CACACE
Institute of Protein Biochemistry and Enzymology, CNR, Via Tioano, 6, 80072 Arco Felice, Naples, Italy

Maori victory

SIR—Whether or not he was summarizing the views expressed in Crosby's *Ecological Imperialism, The Biological Expansion of Europe, 900–1900*, Roy Porter (*Nature* **323**, 587; 1986) shows extreme ignorance when he suggests that the Maori people of Aotearoa (New Zealand) "were stuck in the time-warp of the Stone Age and could not compete with European arms, ploughs and liquor". He also describes them as having been ruined by syphillis, a suggestion that might be equally appropriately applied to the English monarchy.

The Pakeha (European) myths about the land wars of last century have now been exposed. The British were defeated militarily by the Maori in spite of General Cameron's 18,000 men, the largest standing army at that time in the British Empire. The Maori applied lessons learned during their own musket wars to counter the superior fire power of mortars, rockets, howitzers and the 118-pound Armstrong gun. The gunfighter's *pa* (fortress) consisted of a system of trenches, parapets, rifle pits, covered walkways, fire-proofing and bomb shelters. The success of the Maori military strategy was evident at Gate Pa where Cameron suffered his heaviest defeat and concluded that New Zealand could not be taken by a war of attrition.

A more efficient way of getting Maori land was devised when the Maori Land Court was created and a system of individual ownership was installed to break up tribal lands. The war between Maori and Pakeha over the land and its resources was transferred to the arena of Pakeha-created courts, where it continues today.

The judgement that European domination of the globe is some kind of darwinian success story is premature. Maori resistance to imperialism has been so strong that the Pakeha population of Aotearoa has now been co-opted into the struggle. The parliament of New Zealand is about to pass a law that the governments of both the United Kingdom and the United States interpret as banning their navies from our harbours. We invite Alfred Crosby and Roy Porter to come to New Zealand and expand their view of global biopolitics, but we should warn them that they may catch the anti-nuclear bug which is said to be reaching epidemic proportions in this region.

RANGINUI WALKER
Department of Maori Studies,
PETER WILLS
Department of Physics, University of Auckland, Private Bag, Auckland, New Zealand

Pan(da)selectionism

SIR—I have been following with interest the recent correspondence in your pages on evolution, creation science and neo-darwinsim. The letter from T.D. Iles[1], pointing out the regrettable habit of neo-darwinians of making exclusive claims on the interpretation and exposition of evolutionary orthodoxy, was nicely complemented by the Commentary[2] two weeks later on the ancestry of the giant panda. This article, written by a leading (indeed a founding) neo-darwinian, clearly shows the plasticity of neo-darwinian logic.

I do not wish to support either side in the argument over whether the giant panda is more closely related to the raccoon or to the bear (indeed, this is the sort of argument that seems to excite the interest of neo-darwinians in particular: its significance for evolutionary theory in general is less clear), but instead to point out that Mayr's arguments concerning the data on haemoglobin amino-acid sequence differences are circular, self-serving and without experimental proof.

Mayr is concerned on the one hand to prove the irrelevance of the haemoglobin sequence data that show giant pandas to be more closely related to raccoons than to bears, and on the other hand to explain the near identity of the haemoglobin chains of humans and chimpanzees, despite an apparent evolutionary separation of over 10 million years. The explanation offered is, of course, selective pressure on haemoglobin. To quote: "In the case of man and chimpanzee this influence has been conservative, but in the case of the bears it might well have been centrifugal".

What can one say about such *ad hoc* panselectionist arguments that are not supported by biochemical data plausibly relating haemoglobin structure and function to the physiology, habits and behaviour of pandas, raccoons, bears or primates? It is clear that faith in natural selection alone can explain all in these sermons from the pulpit of neo-darwinism. It is surely time to acknowledge that gene and protein sequence data often support theories of molecular evolution in which natural selection of variants is not explanatory[3], or in which common molecular ancestry is disputed[4], and that the use of these data to support either molecular-clock based phylogenetic trees or neo-darwinism selectionist explanations of evolution should be treated with appropriate scepticism.

GREGORY W. WARR
Department of Biochemistry, Medical University of South Carolina, 171 Ashley Avenue, Charleston, North Carolina 29425-2211, USA

1. Iles, T.D. *Nature* **323**, 576 (1986).
2. Mayr, E. *Nature* **323**, 769–771 (1986).
3. Dover, G.A. & Flavell, R.B. *Cell* **38**, 622–623 (1984).
4. Schwabe, C. *Trends biochem. Sci.* **11**, 280–283 (1986).

Solid-state chemistry

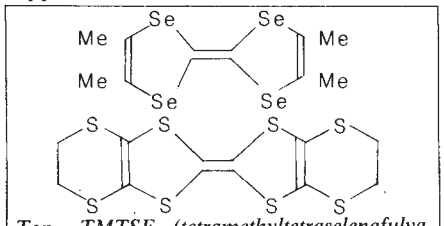
Novel molecular metals

Martin R. Bryce

THE scope of materials that display interesting electrical, magnetic and structural properties at very low temperatures has now been extended by Underhill and co-workers, who on page 547 of this issue[1] describe the first two-dimensional organometallic compound that exists in an integral oxidation state. The study of the synthesis and solid-state properties of such conducting two-dimensional systems will help to elucidate many fundamental phenomena of solid-state chemistry.

Most organic and organometallic solids are electrical insulators, but considerable attention has recently been focused on an ever-increasing number of these solids that exhibit high electrical conductivity and other properties typical of metals. This metallic behaviour is caused by extensive intermolecular overlap of systems rich in π-electrons along highly ordered stacks within the crystal. The materials are highly anisotropic and are termed quasi-one-dimensional molecular metals[2,3].

Because of instabilities inherent in a one-dimensional electron system, metallic behaviour is almost invariably destroyed at low temperature by a Peierls transition when the materials become semiconducting or insulating. (This transition is analogous to the Jahn–Teller symmetry-breaking distortion in molecules.) A major goal of research on molecular metals is the stabilization of the metallic state at very low temperatures. The report by Underhill and colleagues of a new metal–dithiolene complex with an inherently two-dimensional structure, a feature which suppresses the usual Peierls transition and stabilizes the metallic state down to 1.4 K under a pressure of 12 kbar, is therefore of considerable interest.



Top, TMTSF (tetramethyltetraselenafulvalene); bottom, BEDT–TTF (bis(ethylenedithiolo)tetrathiafulvalene).

A significant degree of two-dimensional behaviour has been found previously in molecular metals which are salts of TMTSF and BEDT–TTF (see figure), where X-ray analysis reveals close interstacked selenium–selenium or sulphur–sulphur contacts[4], but two-dimensionality is a new structural feature for highly conducting organometallic salts.

The physical basis of the Peierls transition in molecular metals is a direct consequence of the one-dimensional nature of their crystal structures. The component molecules crystallize in an orientation such that their electronic states mix to form a continuous band along only one crystal axis. The highest energy level within the band that the electrons occupy is termed the Fermi level. At low temperatures, long-range order cannot be sustained and lattice distortions lead to the opening of a band gap at the Fermi level. When this occurs the conducting chain can be considered to stretch in one region and contract in another, so that the conducting electrons become localized with a filled electron band at a lower energy and an empty band at a higher energy. This transition usually occurs at a well-defined temperature (typically between 20 and 100 K) often when the component molecules of the conducting stack dimerize. Above this temperature electrons can be thermally excited across the Peierls gap and a uniformally spaced (metallic) stack is energetically preferred.

In contrast, for a two- or three-dimensional system, where there is also significant coupling between molecules in a plane perpendicular to the stacking axis, it is less favourable, or even impossible, for the lattice to rearrange and open a band gap. The application of pressure to the sample is known to increase the strength of the two-dimensional interactions in TMTSF and BEDT–TTF salts[4].

The discovery by Underhill and colleagues of a complex in an integral oxidation state that nonetheless displays metallic conductivity is also a significant advance. When a complex is in a non-integral oxidation state there is only partial charge on each anion. This means that energy levels will exist close to the Fermi level that are empty and available to receive the electrons required for conduction along the stack. This situation typically arises when there is full charge transfer between donor and acceptor molecules that crystallize in the ratio 1:2 or 2:3, or when there is only partial charge transfer from donor to acceptor that are present in a 1:1 ratio. Conversely, for a complex in an integral oxidation state there is a completely filled band and energy is needed to activate electrons into states where they will be available for conduction. Such materials are typically semiconductors or insulators. In simple terms, movement of electrons along the stack will now necessitate placing two electrons on the same acceptor, with the associated unfavourable coulombic repulsion between these two electrons. □

1. Parker, I.D., Friend, R.H., Clemenson, P.O. & Underhill, A.E. *Nature* **324**, 547 – 549 (1986).
2. Bryce, M.R. & Murphy, L.C. *Nature* **309**, 119–126 (1984).
3. Wudl, F. *Accts Chem. Res.* **17**, 227 – 232 (1984).
4. Williams, J.M. *Prog. inorg. Chem.* **33**, 183 – 220 (1985).

Martin R. Bryce is Lecturer in Organic Chemistry, Durham University, Durham DH1 3LE, UK.

Pattern formation

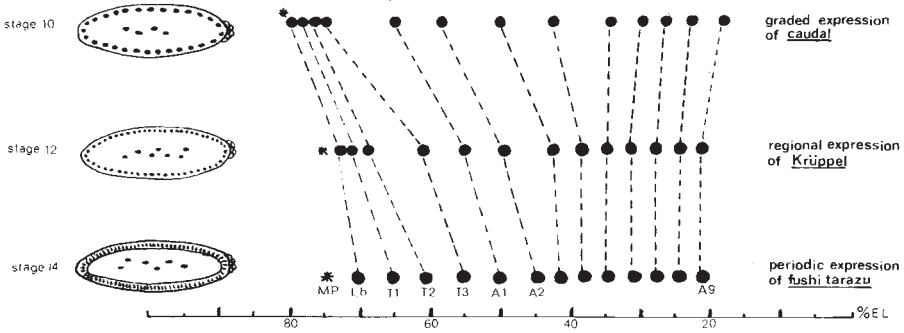
Form and diffusion

Alfonso Martinez-Arias and Philip Ingham

THE question of how cells estimate their location within the body has preoccupied embryologists for the past century, but when posed by Stumpf[1] almost exactly 20 years ago, it seemed to have the beginnings of an answer. The concept of gradients, first explicitly propounded by T.H. Morgan, was entering a new phase of development, and since then it has become almost axiomatic to theories of pattern formation in insects. Yet for all that, direct evidence of the existence of gradients has remained tantalizingly elusive. The recent discovery[2] of the fruitfly (*Drosophila*) gene *caudal* (*cad*), whose transcripts form a concentration gradient in the early embryo, suggested that the search might at last have borne fruit. On page 537 of this issue[3], Macdonald and Struhl report an extensive analysis that argues against a role for *cad* as a primary morphogen. In addition, the detailed characterization of primary determinants of polarity in *Drosophila* by Nüsslein-Volhard and colleagues[4,5] suggests that a reassessment of the gradient paradigm is timely.

Since the beginning of this century, gradients have been invoked as determinants of polarity in a wide variety of organisms[6]. Analysis of polarity perturbations in insect segments had in particular provided support for this view. In the 1960s both Stumpf[1] and Lawrence[7] realized that the postulated gradients of "unknown diffusible substances"[1] could not only determine polarity, as Morgan had suggested as early as 1905, but also "that a particular level of the gradient might determine the pattern differentiated by cells"[7] at any given position. This concept was later embodied in Wolpert's theory[6] of positional information, according to which cells in a morphogenetic field read the local concentration of a morphogen gradient, interpret the information and differentiate accordingly.

Although these authors had been con-



Shift in developmental capacities of the developing blastoderm (adapted from ref. 14). In the stage 10 embryo, when *cad* exhibits a graded distribution, the apparent positions (%EL = egg length) of the segmental primordia (black dots; MP, mouth parts; Lb, labium; T, thoracic; A, abdominal) show a bipartite distribution with clustering towards either pole. The shift in these positions between stages 10 and 14 is consistent with diffusion of signals dependent on *osk* and *bcd*. Their periodic distribution at stage 14 reflects the expression of *fushi tarazu* and other pair-rule genes.

cerned with patterning in secondary fields, the basic concepts of the positional information theory can in principle be applied to the establishment of pattern along the primary body axis during embryogenesis. Indeed, Sander's experiments[8] with insect embryos had already led him to propose a role for gradients in generating the insect body plan. These experiments excluded the possibility that the cells of the early embryos are specified in response to a fixed set of localized determinants. Rather, they encouraged the view that increasing concentrations of morphogens could account for the generation of pattern through the differential activation of genes along the anteroposterior (AP) axis of the embryo[8-11].

At first sight, the *cad* product seems to be a good candidate for such a morphogen. Using antibodies raised against the *cad* protein, Macdonald and Struhl[3] and Mlodzik and Gehring[12] find that, like the transcript, the *cad* protein is initially gradially distributed in the early embryo. The significance of this distribution is, however, diminished by the results of Macdonald and Struhl's extensive genetic analysis of the locus. They have generated mutations that allow them to remove the maternal contribution of the gene from the egg. When such eggs are fertilized by normal sperm, the characteristic early protein distribution is barely detectable yet the embryo develops normally.

Given the disparity between the levels of *cad* protein with and without the maternal contribution, this result argues that the absolute concentration of the gene product at different points along the AP axis does not have any informational content *per se*. Only when the gene is also absent from the zygote is patterning disrupted. Thus it seems it is the localized expression of *cad* in the posterior region of the embryo, rather than its initial graded distribution, which is necessary for normal patterning. Interestingly, the defects observed are predominantly of the 'pair-rule' type. Moreover, they are correlated with alterations in the pattern of expression of the pair-rule gene *fushi tarazu* similar to those caused by the 'knirps-grandchildless' group of mutations[11,13]. This suggests that *cad* is involved in the transfer of information between the primary determinants of pattern and zygotic segmentation genes.

Some of these primary determinants have recently been identified in a series of elegant analyses by Nüsslein-Volhard and colleagues. Starting from the premise that the primary determinants of the body plan must be deposited in the egg during oogenesis, her group has isolated maternally acting mutations which disrupt patterning along the AP axis. One of the genes so identified, *bicoid* (*bcd*), is required for the normal development of the anterior, whereas another, *oskar* (*osk*), is necessary for normal abdominal development.

By carrying out various cytoplasmic transplantations, Nüsslein-Volhard and colleagues show that activities dependent on the two genes are initially localized at the anterior and posterior poles of the embryo, respectively. In both cases, the mutant phenotype can be repaired by transplantation of wild-type cytoplasm form the appropriate pole. In the case of *osk*, the failure to generate the abdominal segments can only be rescued if the posterior polar material is transplanted to the prospective abdominal region of the embryo and not to the posterior pole. Normal cytoplasm from the abdominal region does not, however, have rescuing activity. This suggests that the material localized in the posterior pole acts as a source which propagates a molecular signal which in some way is transferred to the abdominal region during the early cleavage stages.

Similarly, *bcd* activity seems to be localized at high levels in the anterior pole of the egg. Transplantation of this anterior polar material not only rescues the *bcd* defect but also can induce anterior structure at any position along the AP axis. The concentration of this activity decreases both with distance from the pole and with time, suggesting a dynamic gradient of *bcd* activity in the anterior part of the embryo. The status of both gene products as primary morphogens is reinforced by the finding that their transplantation to ectopic sites in the embryo can also induce changes in polarity, its most fundamental property.

Despite this compelling evidence for the existence of two sources of diffusible influences emanating from either pole, Nüsslein-Volhard's group does not interpret its data within the framework of the clasical gradient paradigm. Rather than opposing gradients of *osk* and *bcd* products specifying unique values along the AP axis, both genes are thought to supply more coarse-grained information that would be responsible for the region-specific activation of the zygotic gap genes.

In the light of these results, it is interesting to re-examine some of the earlier classical embryological manipulations. In 1977 Vogel[14] reported a series of ligature experiments in *Drosophila* analogous to those performed by Sander on other insects. On the basis of these, he argued against the generation of the body plan through interacting gradients of morphogens, but instead proposed a regionalization of the embryo consistent with the results discussed above (see figure).

Remarkably, the emerging picture seems not inconsistent with the vision of Driesch[15] who at the turn of this century proposed that "development starts with a few ordered manifoldnesses, but the new manifoldnesses create by interactions, and these are able ... to provoke new differences and so on. With each effect immediately a new cause is provided and the possibility of new specific action".

Although all the new results prompt a re-evaluation of the role and mechanism of action of gradients in primary pattern formation, they do not have any direct bearing on the same aspects of secondary fields originally addressed by Stumpf and by Lawrence. Similar experimental approaches will no doubt sharpen our understanding of the latter. □

1. Stumpf, H. *Nature* **313**, 430 – 431 (1966).
2. Mlodzik, M., Fjose, A. & Gehring, W.J. *EMBO J.* **4**, 2961 – 2060 (1985).
3. Macdonald, P.M. & Struhl, G. *Nature* 324, 537 – 545 (1986).
4. Lehman, R. & Nüsslein-Volhard, C. *Cell* **47**, 141 – 152 (1986).
5. Frohnhöfer, H.G. & Nüsslein-Volhard, C. *Nature* **324**, 120 – 125 (1986).
6. Wolpert, L. in *A History of Embryology* (eds Horder, T.J., Witkowski, J.A. & Wiley, C.C.) 347 – 362 (Cambridge Univ. Press, 1986).
7. Lawrence, P.A. *J. exp. Biol* **44**, 602 – 620 (1966).
8. Sander, K. *Adv. insect Physiol.* **12**, 125 – 238 (1966).
9. Lewis, E.B. *Nature* **276**, 565 – 570 (1978).
10. Nüsslein-Volhard, C. *Wilhelm Roux Arch. dev. Biol.* **183**, 247 – 268 (1977).
11. Schüpbach, T. & Wieschaus, E. *Wilhelm Roux Arch. dev. Biol.* **195**, 302 – 317 (1986).
12. Mlodzik, M. & Gehring, W.J. *Cell* (in the press).
13. Carrol, S.B. *et al. Nature* **323**, 278 – 280 (1986).
14. Vogel, O. *Wilhelm Roux Arch. dev. Biol.* **182**, 9 – 32 (1977).
15. Stern, C. *Caryologia* **6**, 355 – 369 (1954).

Alfonso Martinez-Arias is at the MRC Laboratory of Molecular Biology, Cambridge CB2 2QH and Philip Ingham is at the ICRF Developmental Biology Unit, Department of Zoology, Oxford OX1 3PS, UK.

Plasma physics

Nuclear fusion comes closer

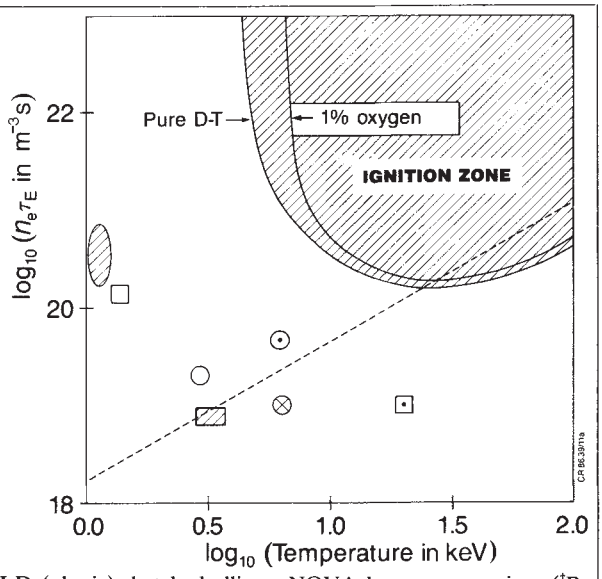
R.S. Pease

A NEW initiative in international cooperation among the nuclear fusion research community, which may lead to a world engineering test reactor, was announced at a recent meeting* (see below). But the dominant scientific feature of the conference was the experimental results from four large tokamak installations, TFTR (Princeton), D-III-D (San Diego), JET (Euratom, Culham) and JT-60 (JAERI, Japan). All these tokamaks, toroidal electric discharges stabilized by strong magnetic fields, have discharge currents in the range 1 – 5 MA and additional heating in the power range 1 – 20 MW, and are aimed to investigate magnetic confinement of high-temperature hydrogen or deuterium in conditions close to those required for net energy production in a deuterium–tritium (D–T) mixture (see my recent News and Views article, *Nature* **324,** 111; 1986). TFTR and JET, however, are not now planned to operate in D–T before 1990.

The results achieved in simple terms of temperature and the critical thermal insulation parameter $n\tau_E$, where n is the ion density and τ_E the energy confinement time, are summarized in the figure, and are compared with the values needed to obtain a self-sustained thermonuclear reaction in a 50 per cent D–T mixture. The highest ion temperature (T_i) reported, achieved with the strongest additional heating and at low densities ($\sim 10^{19}$ m^{-3}), is about 20 keV; the best thermal insulation, achieved generally with low or no additional heating and at the highest densities, is about 1.5×10^{20} m^{-3} s, that is, just at the ignition threshold.

The best combination, represented by the triple product $n\tau_E T_i$, is about 2×10^{20} m^{-3} s keV obtained both in JET and in TFTR, and is about a factor of five above that reported at the previous conference in 1984; it is still about a factor of ten short of the minimum 3×10^{21} m^{-3} s keV required for self-sustained thermonuclear reaction in D–T. Up to 10 kW of actual fusion power has been achieved in deuterium alone, which corresponds to more than 1 MW in D–T. Typically, such conditions can be sustained for several seconds in JET, the largest tokamak.

After allowing for the various differing features of these machines, and for the initial nature of the experiments in D-III-D and JT-60, the major experimental results are all in substantial agreement. Moreover, the results from these large tokamaks are supported and supplemented by a wide range of heating and confinement experiments conducted on numerous smaller tokamak installations such as the T-10 (Kurchatov), the DITE and CLEO assemblies (Culham), the HL-1 tokamak in Leshan (China) and high-field tokamaks such as ALCATOR C. Generally, all the additional heating methods raise the plasma temperature markedly, have high efficiences and are understood. The main phenomenon observed is that the best confinement times are found in high-density operation (Alcator scaling) and in the largest machines. But the scaling with temperature and magnetic field is otherwise not very marked.



Experimental achievements of ion temperature T_i (generally, the central ion temperature) and density–energy confinement time product ($n\tau_E$) for the four large tokamaks† and the NOVA laser compression experiment (E. Storm, Livermore). These results are compared with the theoretical values needed for ignition in a D–T plasma (see my review, *Phil. Trans. R. Soc.*, in the press). Below the dotted line, the ion and electron temperatures can be significantly different. Experimental results of: ○, JET (ohmic); ⊙, JET (X-point operation); □, TFTR (ohmic with pellet injection); ⊡, TFTR (best temperature); ⊗, JT-60 (neutral beam heating); hatched rectangle, D-III-D (ohmic); hatched ellipse, NOVA laser compression. (†R. Hawryluk *et al.* and S. Milora *et al.*, Princeton; I. Tanga, JET; P.-H. Rebut, JET; M. Yoshikawa, JAERI; and J. Luxon, San Diego.)

The chief lacuna is that the physical mechanisms underlying the observed energy confinement time remain unidentified. Ideally, if the heat transport across the magnetic fields were caused by only ion–ion and ion–electron encounters, then the observed value of τ_E (0.1 – 1 s for large installations) should be about ten times larger; and the corresponding values of $n\tau_E$ should increase as the square of the current. In this sense, all four installations have the potential to achieve substantially improved confinement. There is growing evidence that this can be done.

Pellets

One practical route to better values of $n\tau_E$ is to increase the number density, for example by injecting high-speed (about 1 km s^{-1}) pellets of frozen deuterium. This approach is used particularly by the TFTR group and, because τ_E at least does not fall with density (in fact it usually increases), it has met with success which should improve as better pellet injectors are installed. Another approach is to adjust the field configuration in the outer regions (X-point operation). This technique, discovered experimentally on the German tokamak ASDEX (Wagner, F. *Phys. Rev. Lett.* **49,** 1408; 1982) has met with success in the first reported experiments on JET and on D-III-D. The adjustment eliminates the further degradation of confinement time which normally accompanies the additional heating. The radial profile of the current density (in the minor cross-section) can be more generally altered by non-inductive current drive: up to 1.7 MA was generated by travelling wave radio frequency (2 GHz) fields in JT-60; but the favourable effects on τ_E are restricted so far to low densities. Lastly, under a number of conditions, the thermal insulation is observed to show transient improvements localized near the plasma centre, which leads to a strong rise of temperature lasting about a second, followed by a collapse. This, again, suggests that long-lasting improved confinement can be obtained.

Theoretical studies of the transport problem, stimulated by new, well-diagnosed experimental data, are growing. No major instability is, it seems, predicted that could inevitably preclude achieving the required thermal insulation. Those that are predicted depend noticeably on the detailed profiles of density, temperature and current, and are governed by the detailed velocity distribution of the ions or electrons and their collision rates. None, so far as was reported at the meeting, gives a convincing numerical explanation of the observed heat losses. It is quite probable — indeed almost certain — that different mechanisms predominate

*Biennial IAEA nuclear fusion research meeting at Kyoto, Japan, 13–20 November 1986.

in different regions of the discharge, so that reducing complex theory to a simple confinement time is no easy task. An encouraging development of the experiments is that the impurity levels (~ 1 per cent of O and C) are insufficient to produce a crippling radiation loss.

Taken together, the experiments on the large tokamaks reported substantial improvements in temperature and confinement time arising from the larger dimensions of the new apparatus, from the pellet injection and the powerful additional heating. There is a broad empirical basis for expecting further improvement in these parameters to be achieved on these installations, and there are good prospects that the extra experimental diagnostics and the theoretical efforts will lead to the required understanding of the transport process. The question of whether even the largest of the installations, JET, can reach partial thermonuclear ignition in D–T fusion is not yet resolved, but tokamak research is clearly flourishing.

From the experiments in other magnetic confinement systems reported at the meeting, progress is largely resulting from improvements to already established assemblies. For example, a several-fold improvement in confinement time was reported from the HBTX reverse-field pinch experiment at Culham, resulting from careful elimination of defects in the magnetic field system. However, the confinement time and temperature in these systems are below those achieved by the strong field stabilization of the tokamak; and are further from the theoretical ideal. Both the Japanese (Heliotron E) and German (Wendlestein) groups are operating stellarators using essentially the same apparatus as before. Big advances can be expected only when apparatus designed to explore more advanced conditions is brought into operation, for example, new stellarators under construction in Garching, in Canberra and in Oak Ridge, and the reverse-field pinch RFX at Padua.

Lasers

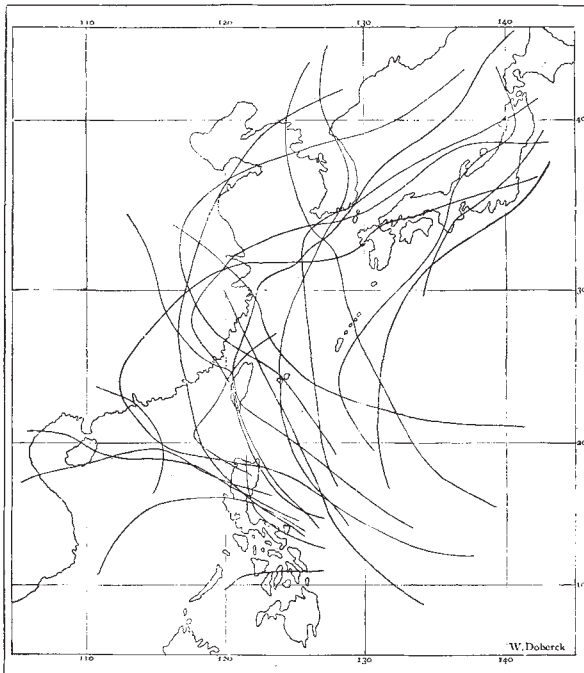
Competition for the tokamaks was also reported from inertial confinement experiments, mainly from laser installations in the United States, the Soviet Union, Japan and Europe, but especially from the major new laser installation NOVA at Lawrence Livermore Laboratory, United States and Gekko 12 at Osaka with 70 and 30 kJ output pulses from neodymium glass and phosphate glass, respectively. Temperatures of up to 10 keV and densities of up to 50 times liquid were reported. Using neutron pulse duration as a measure of τ, the Livermore group reported 2–4 × 10^{20} m^{-3} s as a value $n\tau_E$ at 1 keV, but it was unable to give details of the derivation because its work is classified. The direct implosion of D–T contained in thin-walled glass microspheres, where time-profiled radiation pulses were used to mitigate Raleigh–Taylor instabilities, were reported to give the best promise of a net useful energy yield. Compression by hohlraum radiation was reported, as also were new target designs using CH_2 foam to stabilize the implosion. Both groups reported neutron yields at record values of 10^{13} neutrons per pulse, corresponding to about 10^{-5} burn up of the pellet fuel, and a 2 × 10^{-3} energy yield compared with the absorbed laser energy. The laser approach is hampered by the low energy efficiency of lasers; the first engineering tests of the pulsed beam fusion assembly (PBFAII) at Sandia laboratory are therefore important because the charged-particle beams offer the prospect of much higher efficiency. The formidable technology problems of 30-MV pulsed transmission lines for nanosecond operation are being overcome; a liquid-lithium ion source based on a dense array of electrostatically induced Taylor cones (J. Vandevender, Sandia) is being developed to provide ion emission intensities of 5 kA cm^{-2}.

The impression given by these and about 20 papers from smaller installations in France, United States, Britain, China, Canada and Japan is that inertial confinement of high-temperature dense matter is a fast-developing area of physics with many new ideas and results. The standard of experiments led by the new US and Japanese facilities is very high.

In the sessions devoted to envisaged fusion reactors and reactor technology, one new feature was the first design of a fusion reactor based on the muon catalysed fusion reaction (S. Elezier *et al.* Israel and Texas), muon catalysis being the dark horse of the fusion race. An outstanding feature of the engineering is the successful operation by Kyushu University, Japan of the first superconducting tokamak operating with $NbSn_3$ coils. The first experiments with plasma use up to 170 kA and the full fields of 11 tesla (at the coils). There is agreement between experiment and theory on the important issue of shielding the superconducting winding from magnetic transients generated by violent plasma current disruption. The engineering problems of the difficult $NbSn_3$ coils, which have delayed completion of the larger Soviet T-15 tokamak, seem to have been overcome in this relatively small scale (major radius 0.8 m) Triam apparatus.

The generally bullish outlook on the potential for progress of the magnetic confinement approach to fusion was emphasized by the announcement (J.F. Clarke, Artsimovitch memorial lecture, 13 November, 1986) of a new proposal for international cooperation. Following discussions at the Reykjavik summit meeting, the US government on 30 October transmitted a formal proposal to the Economic Summit nations and to the Soviet Union. According to Clarke the proposal is to join together to prepare a conceptual design report (CDR) for an engineering test reactor. The proposal indicates that the United States is prepared to commit sufficient resources to accomplish this task within three years. If accepted, Clarke believes that this proposal could lead to a major commitment of international resources on preparing for an engineering the most widely accepted element of a common fusion plan. □

R.S. Pease is Programme Director for Fusion at the UK Atomic Energy Authority, Culham Laboratory, Abingdon, Oxfordshire OX14 3DB, UK.



From *Nature* **35**, 135; 9 December 1886.

100 years ago

THE LAW OF STORMS IN THE EASTERN SEAS

PRINCIPAL typhoons 1884–1885. Nearly all the typhoons appear to have their origin east or south-east of the high-pressure area that covers the northern Pacific in the summer season, which part of the ocean is characterised by high sea-surface temperature. Typhoons are sometimes formed in the China Sea, but then they seldom develop much energy, as they usually move quickly northwards and enter the mainland of China or Formosa. Owing to their small dimensions they are easily avoided by such ships as may fall in with them. If, however, a typhoon of this kind passes northwards up through the Formosa Channel, it soon becomes as formidable as any of those that originate in the tropical Pacific.

Biological diversity

How many species are there?

Robert M. May

ONE of the first things that an extraterrestrial might ask about the planet Earth is how many species are on it. If this extraterrestrial had only our own current knowledge to rely on, the answer would be astonishingly vague: somewhere between 1.5 and 30 million species of plants and animals.

At one end of this spectrum, the number of genera and species of mammals, birds and other large animals are fairly accurately known. As summarized by Diamond[1] in a News and Views article last year, new mammals and birds continue to be discovered (at the rate of about one genus of mammal and three species of birds per year), mainly in the tropics. When we consider the insects and other smaller creatures, however, great uncertainties arise.

Until a few years ago, the uncertainties seemed somewhat less than they do today, and the total number of species was widely believed to lie in the range from 1.5 to 3–5 million. This estimate was obtained roughly as follows. For the species of mammals, birds and other larger animals that are relatively well enumerated, there are roughly two to three times more species in tropical regions than in temperate ones. The total number of species actually named and catalogued is just over one million. Most of these are insects. Indeed, to a good approximation, all species are insects! But most insects that have actually been named and taxonomically classifed are from temperate zones. Thus, if the ratio of numbers of tropical to temperate species is the same for insects as for mammals and birds, we may expect there to be something like two or three yet-unnamed species of tropical insects for every one named temperate species. Hence the overall crude estimate of a total of roughly 3 times the number currently classified, or around 3–5 million.

The dramatic upward revision of this estimate results from Erwin's careful studies of the insect fauna in the canopy of tropical trees[2-4]. Erwin used an insecticidal fog (generated by a machine hoisted into the canopy) to 'knock down' the canopy insects; because of their inaccessibility there have been few previous studies of the insect fauna in tropical canopies. Erwin's studies suggest that most tropical arthropod species in fact live in the tree tops, which is not so surprising as this is where there is most sunshine as well as most green leaves, fruits and flowers.

Erwin first collected the arthropods from 19 *Luehea seemannii* trees in Panama over three seasons[2,3]. This tree, a medium-sized evergreen of seasonal tropical forests, has an open canopy and large and well-spaced leaves. The trees chosen for sampling had few epiphytes or lianas, and yielded more than 1,100 species of canopy-dwelling beetles (including weevils), distributed among the categories of herbivore, predator, fungivore and scavenger as shown in the table. To use this information as a basis for estimating the total number of insect species in the tropics, one needs to know what fraction of the fauna are specific to the particular three species or genus under study. Unfortunately, as Erwin emphasizes[3], "No data are available with which to judge the proportion of host-specific arthropods [on trees] per trophic group anywhere, let along the tropics. So, conservatively, I allowed 20 per cent of the *Luehea* herbivorous beetles to be host-specific (that is, they must use this tree species in some way for successful reproduction); 5 per cent of the predators (that is, tied to one or more of the host-specific herbivores); 10 per cent of the fungivores (that is, tied to fungus associated only with this tree); and 5 per cent of the scavengers (that is, are associated in some way with only the tree or with the other three trophic groups)". The results summarized in the table (which is slightly modified from Erwin's work in that I thought it inappropriate to go beyond two-figure accuracy in the last column), suggest about 160 host-specific canopy beetles per genus-group of trees.

Estimated numbers of host-specific canopy beetles on *Luehea seemannii*, classified into trophic group (from ref. 3).

Trophic group	Number of species	Estimated fraction host-specific (%)	Estimated number of host-specific species
Herbivores	682	20	140
Predators	296	5	15
Fungivores	69	10	7
Scavengers	96	5	5
Total	1,100+		160

Several other assumptions are needed to extrapolate from the estimate of 160 host-specific species of canopy beetles per tree to 30 million species in total[3,4]. Slightly simplified, the argument runs as follows. First, Erwin notes that beetles represent roughly 40 per cent of all known arthropod species, leading to an estimate of 400 canopy arthropod species per three species. Next, he suggests the canopy fauna is at least twice as rich as the forest floor fauna, and is composed mainly of different species; this increases the estimate to around 600 arthropod species that are specific to each genus-group of tropical tree. Citing an estimate of 50,000 species of tropical trees[5], Erwin thus arrives at the possibility that there are 30 million tropical arthropods in total.

It is easy to cavil at each step of this chain or argument. For instance, the fact that 40 per cent of taxonomically classified arthropod species are beetles is of doubtful relevance if in truth essentially all arthropods are unclassified tropical canopy dwellers; what we need to know is what fraction of this canopy fauna consists of beetles. Furthermore, the overall estimate depends almost linearly on the necessarily arbitrary assumption that 20 per cent of the herbivorous beetles are specific to a genus-group of tree; changing this number to 10 per cent would halve the global estimate to 15 million species. Nor is it clear to me that each of the 50,000 species of tropical trees constitutes a 'genus-group' in the sense used by Erwin. These reservations do not, however, detract from my admiration of his work, which has advanced us to the point where we can begin systematic investigation of each of the links in his chain of argument; in comparison, the earlier estimate of 3–5 million is pure hand-waving.

There is a different argument, altogether independent of any of those I have discussed so far, which suggests the total number of species exceeds 10 million. A plot of the number of species as a function of the physical size of the constituent individuals (on a log–log plot) gives an almost linear regression line, with a slope around −2, for species with characteristic linear dimensions above 1 cm or so[6]. That is, for the relatively well-catalogued larger species a tenfold reduction in linear dimensions results in roughly a hundredfold more species; there are roughly 100 as many species of characteristic size 10 cm than of size 1 m. But as we move below 1 cm or so, the number of classified species falls further and further below the regression line that fits larger size categories[6]. Were this regression extrapolated down to around 0.5 mm, we would increase the species total from the known 1 million or so to more than 10 million. As we lack a fundamental understanding of the size-species relation itself, there is no reason to expect such a simple extrapolation to estimate accurately the number of unclassified smaller species, but it is interesting that the number is not inconsistent with

Erwin's more biologically based estimate.

Moving on to smaller organisms, we have to consider protozoa, bacteria and viruses. Some would argue that the species concept breaks down when applied to viruses and bacteria, with their recombining plasmids and the like. Others observe that the small number of species of 'microparasites' currently classified are disproportionately pathogens of *Homo sapiens,* and that a highly conservative estimate would be one such microparasite species particular to each species of invertebrate or vertebrate animal or plant. This would double the total species count at a stroke. Whether anyone cares for the preservation of these microparasite species is another question. A touching note in the Smithsonian Museum of Natural History records the passing in 1914 of Martha, the last passenger pigeon; no-one wept for the passing of the last wild smallpox virus. Many popular ecology books explain how predators help the prey species by culling diseased individuals, but I have yet to see the suggestion that microparasites are beneficial in supplying weakened prey to sustain predator populations.

In between the furries and featheries that excite compassion and the viral, bacterial and protozoan pathogens that we try to eliminate lie the insects, to which both taxonomists and conservationists have been relatively indifferent. Erwin's work is important in focusing concern on this huge class of animals that has been neglected for too long. □

1. Diamond, J.M. *Nature* **315**, 538 (1985).
2. Erwin, T.L. & Scott, J.C. *Coleopt. Bull.* **34**, 305 (1980).
3. Erwin, T.L. *Coleopt. Bull.* **36**, 74 (1982).
4. Erwin, T.L. *Bull. Ent. Soc. Am.* **2**, 14 (1983).
5. May, R.M. in *Diversity of Insect Faunas* (eds Mound, L.A. & Waloff, N.) 188–204 (Blackwell, Oxford, 1978).

Robert M. May is Class of 1877 Professor of Zoology at Princeton University, Princeton, New Jersey 08544, USA.

Halite rheology

Salt with a pinch of water

Malcolm K. Jenyon

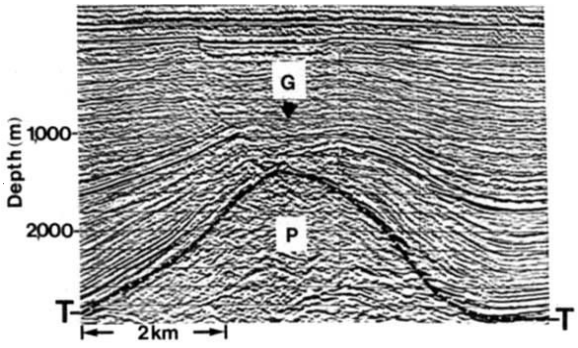
THE connection between the rheological behaviour of the mineral halite (NaCl) in the presence of very small quantities of free water and radioactive waste disposal may at first seem obscure. There is a strong link, however, some aspects of which are described by Urai *et al.*[1] on page 554 of this issue. Adding to previous work[2,3], these authors go some way towards elucidating the marked effects on halite rheology of the very small quantities of brine inherent in a polycrystalline salt mass. The investigations have important implications for the determination of long-term stability in any rock salt body that is being considered as a repository of radioactive waste or for material storage. Their findings are also of importance in the wider field of salt tectonics.

Natural deposits of rock salt, the main constituent of which is halite, usually occur as members of evaporite cycles in stratified layers or as the thickened 'pillows' and piercement structure (diapirs) resulting from plastic deformation of the source salt layer. These salt masses are attractive as sites for material storage or waste disposal for two main reasons: first, rock salt at depth is impermeable to the types of material involved; and second, by solution mining it is relatively easy to create storage caverns in the salt of sufficient volume to be economically viable. Statistics of usage are not easy to come by (particularly in the case of radioactive waste), but as an example of use for hydrocarbon storage, Dreyer[4] mentions that about 140 caverns (most being in salt bodies) with a total volume of 22 million cubic metres were either completed or under construction in the Federal Republic of Germany in 1982.

In the case of radioactive waste disposal, it is clear that determination of the long-term stability of a salt cavern is a priority for reasons of safety, needing a wide range of investigations to study the tectonic framework of the area (including the fault systems and local seismicity) and to measure the stability of the bedrock beneath the salt; that of the groundwater system; the permeabilities of salt and 'country rock'; the geothermal gradient; the internal integrity of the salt body itself; and the stress patterns and movement history of the salt body (see figure).

The question of the movement history of the salt body is important in assessing the long-term stability of any cavern that is constructed; however, the rheological behaviour of halite is still imperfectly understood. It is known that over a range of temperatures far below its melting point of about 800 °C, halite deforms plastically by intragranular dislocation and gliding, and probably by diffusive mass transfer (pressure solution) through thin, intergranular aqueous films. Increasing the temperature reduces the yield strength of halite, as does the application of differential pressure, although increasing the confining pressure seems to have a much smaller effect. The presence of free water also reduces the yield strength of halite; however, as a polycrystalline halite mass buried to depths greater than about 300 metres is thought to be impermeable (see, for example, ref. 5), the effect of free water on plastic deformation in buried salt masses has been little studied.

The work of Urai *et al.*[1] shows that water need not enter a salt body from outside for the weakening effect of free water to be a factor. The authors demonstrate that even the minute quantities of inherent brine in a salt body (that contained in fluid inclusions in the halite crystals, and particularly in intergranular films) may have profound effects on the rheology of salt. Of particular importance are indications that, whereas at higher strain rates the deformation mode is that of dislocation/gliding, at very low strain rates there is a transition to pressure solution, by grain boundary diffusion with dynamic recrystallization along the boundaries. Also, lattice diffusion can occur, with water molecules passing through the NaCl lattice, and this, together with the brine inclusions



Seismic section through a salt pillow (P) produced by plastic deformation of the salt source layer, the top of which is indicated (T). The overburden above T consists of Mesozoic and Tertiary clastic sediments. At G a collapse-graben is seen, resulting from subsidence of the overlying sediments into the incipient void following dissolution and removal of salt by undersaturated water circulation in overburden fractures above the pillow crest, and in the crestal zone of the salt feature. Thickened salt bodies of this type, and the diapiric (piercement) structures into which they frequently evolve, may be considered as sites for cavern construction and the storage/disposal of various materials including radioactive waste. Scales are approximate; vertical scale exaggerated approximately 2.5:1.

1. Urai, J.L., Spiers, C.J., Zwart, H.J. & Lister, G.S. *Nature* **324**, 554 (1986).
2. Spiers, C.J., Urai, J.L., Lister, G.S., Boland, J.N. & Zwart, H.J. *European Atomic Energy Commnity Contract WAS-153-80-7-N(N)* (Inst. Earth Sci., Univ. Utrecht, in the press).
3. Spiers, C.J., Urai, J.L. & Lister, G.S. in *Proc. 2nd Conf. Mechanical Behaviour of Salt 1984* Hannover (in the press).
4. Dreyer, W. *Underground Storage of Oil and Gas in Salt Deosits and Other Non-Hard Rocks* (Halsted, New York, 1982).
5. Baar, C.A. *Applied Salt-Rock Mechanics 1* (Elsevier, Amerstam, 1977).
6. Odé, H. *Am. As. Petrol. Geol. Mem.* **8**, 70 (1968).

available within halite crystals, may affect the dynamics of the dislocation mode of deformation.

The findings go a long way towards explaining apparent inconsistencies (based on current ideas) in salt deformation at different depths and in different areas of some salt basins. They may also explain unusual phenomena such as the marked increase in flow rate during the wet season of the rock salt under normal temperature and pressure at the surface in the emergent Hormuz salt diapirs of Iran, and also the Joffé effect, in which a halite crystal exhibits a marked increase in ductility when deformed under fresh water[6].

Plainly, the results obtained so far demand that attention is paid to the amount and distribution of inherent brine in the rock salt members of evaporite sequences in any basin, to make an effective analysis of the causes and results of plastic deformation. The message that low-stress conditions acting over long periods can produce significant plastic deformation in salt rock will have a considerable impact on our ideas about salt deformation in general. □

Malcolm K. Jenyon is Manager, Interpretation, at Seismograph Service (England) Limited, Holwood, Keston, Kent BR2 6HD, UK.

Phylogeny

The German 'ostrich' and the molecular clock

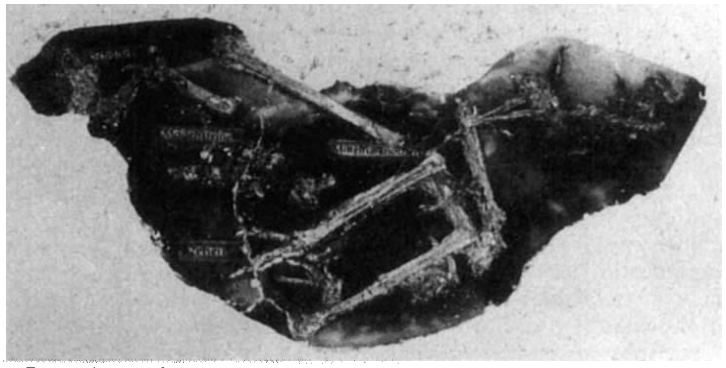
Michael E. Howgate

ONE of the major problems with molecular clocks is that they need to be located in an absolute time frame determined by the fossil record. Undoubtedly the best such clock available at the moment is the DNA–DNA hybridization clock of Sibley and Ahlquist[1], but Peter Houde on page 563 of this issue[2] argues that the DNA–DNA clock needs to be reset.

This molecular clock takes as one of its reference points the divergence of ostrich DNA from rhea DNA. This divergence has been timed from the break-up of the super-continent Gondwanaland which isolated the supposedly already ground-dwelling ratites (ostrich, rhea, emu, cassowaries and kiwi) on the different parts of the fragmenting supercontinent during the late Mesozoic. The palaeontological basis for such a reference point has always been weak as fossils ascribable to the modern genera of the ratites have not been found in strata lower than the Miocene. So it is purely speculative that the ratites were grounded before the break-up of the Gondwanaland supercontinent.

The ratites are a rather odd assemblage of ground-dwelling birds confined to the southern continents and with a basic overall similarity due to their adoption of a flightless condition. What separates them from most other birds, even some which have become flightless, is the relatively primitive, reptilian (palaeognathous) structure of the palate and the presence of rhamphothecal grooves in the mandible. The ratite paletal structure, which is shared with the tinamous (themselves poor fliers) of South and Central America, is intermediate between the palate of archosaurian reptiles and that seen in all other birds — the neognathous condition. This 'primitive' avian palate involves a characteristic arrangement of skull bones, with comparatively large vomers and palatines, the latter being firmly united to the pterygoids producing a less flexible skull than that observed in neognathous birds.

The current consensus of palaeontological and biochemical data indicates that ratites and tinamous are a monophyletic assemblage, although whether they arose from a neognathous ancestor or are primitively palaeognathous is still a matter of contention. Olson[3] outlines five possibilities to account for the palaeognathous condition of these birds: (1) that they are monophyletically derived from neognathous birds; (2) that such a derivation is due to neotony; (3) that their derivation is completely separate from the neognathous birds, with the different ratites evolving from a 'proto-tinamou'; (4) some living palaeognathes are derived from a primitively palaeognathous stock, whereas others are derived from one or more neognathous stocks; and (5) all living palaeognathes are derived from different neognathous stocks.



Boney issues: fossil *Palaeotis weigelti*, from the Geiseltal museum, Halle.

Houde now suggests that an assemblage of palaeognathous birds with well-developed wings and pectoral girdles — the *Linothis*-cohort — are the direct ancestors of all living ratites and evolved distinctive features of the ratite palate and bone histology during the Palaeocene and early Eocene in Europe and North America. This confirms the third of Olson's suggestions except that according to Houde the ratites diversified from a post-tinamous rather than a pre-tinamous condition. Houde further hypothesizes that a bird that retained the power of flight at this time must have flown to New Zealand to give rise to the extant kiwis, which retain some of the primitive characteristics of *Linorthis* that have been lost by other more derived ratites.

Meanwhile in Europe the *Linorthis*-cohort gave rise to an ostrich-grade palaeognath *Palaeotis weigelti* during the middle Eocene (see figure). Thus, ostriches appear to have evolved in Europe and subsequently emigrated overland to Africa at any time from the late Eocene onwards.

If the picture envisaged by Houde is correct and flying ratite ancestors inhabited North America and Europe during the Palaeocene and Lower Eocene and gave rise to an ostrich-grade descendant in the middle Eocene in Europe, long after the break-up of Gondwanaland, then the datum on which DNA–DNA hybridization data is tied in to the fossil record is out by tens of millions of years.

What may be true for ostriches may not apply to the other ratites, and although Houde does not mention it he must hypothesize an almost world-wide distribution for his *Linorthis*-cohort to explain the distribution of the ratites, and he does suggest they must have flown to New Zealand. The lack of fossil evidence for *Linorthis et al.* in the southern continents may only be Northern Hemisphere collecting bias and misidentification of the sort which plagues avian palaeontology. How else is one to explain the presence of the presumed flightless palaeognath *Diagenornis fragilis* from the Palaeocene of Brazil? The problem of the fossil record, however , does not conflict with his critique of the DNA–DNA clock; it is clear the calibration will require further scrutiny. □

1. Sibley, C.G. & Alquist, J.E. in *Evolution Today, Proc. Second Int. Congr. Syst. evol. Biol.* (eds Scudder, G.G.E. & Reveal, J.L.) 301 – 335 (Carnegie-Mellon Univ.).
2. Houde, P. *Nature* **324**, 563–565 (1986).
3. Olson, S.L. in *Avian Biology* vol. 8, 80 – 238 (Academic, New York, 1985).

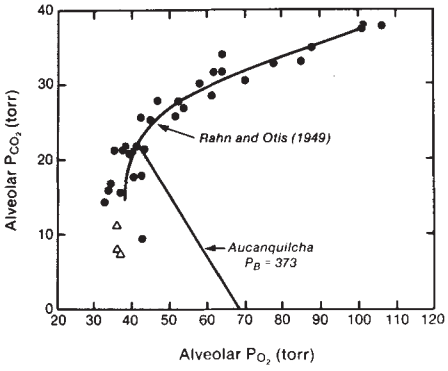
Michael E. Howgate is in the Department of Zoology, University College London, London WC1E 6BT.

Highest inhabitants in the world

SIR—In 1935, members of the International High Altitude Expedition went to the Aucanquilcha sulphur mine in north Chile (21°12′S, 68°28′W) to study their own physiological adaptations to extreme altitude, and also to make some measurements on the local residents[1]. They described a group of miners who worked at the mine, which was said to be at an altitude of 5,800 m, but lived at an altitude of 4,340 m, preferring to climb 500 m on foot every day, which took about an hour and a half, rather than live at the extreme altitude of the mine.

The fact that the miners did not live at the mine itself has often been cited as evidence that people cannot live that high. Furthermore when eight physiologists on the Himalayan Scientific and Mountaineering Expedition of 1960–1961, under the scientific leadership of L.G.C.E. Pugh, spent up to 100 days in a prefabricated hut at an altitude of 5,800 m where the barometric pressure averaged about 380 torr, they seemed to confirm the view. In spite of reasonably good living conditions, ample food, and a healthy exercise programme, they steadily lost weight at the rate of 0.5–1.5 kg per week[2].

It is remarkable that although the Aucanquilcha miners have often been cited as the highest inhabitants of the world[3,4], and therefore are of considerable scientific interest, there is no report of them being visited by physiologists since 1935. Last year, however, I had the opportunity to visit the mine with the Chilean physiologist, Raimundo Santolaya, and was astonished to find that four caretakers were living at the mine itself in a galvanized iron hut. One man had been living there for two years when I met him, but most of the caretakers live there for shorter periods and they all descend by truck to Amincha (4,200 m) on Sundays to play football. Conversations with the manager of the mine made it clear that only a small subset of miners are able to tolerate the altitude. Nevertheless it is clear that some people can live at this altitude for indefinite periods of time. The miners are mostly Bolivian natives who normally reside between 2,000 and 4,000 m.



Oxygen–carbon dioxide diagram showing alveolar gas composition of both permanent residents and lowlanders acclimatized to high altitude. Most of the data were collated by Rahn and Otis[6]. The three triangles bottom left were obtained at extreme altitudes on Mount Everest[7]. Diagonal line shows respiratory exchange ratio of 0.85 for barometric pressure of 372 torr and predicts that the inhabitants at 5,950 m will have alveolar P_{O_2} and P_{CO_2} values of approximately 42 and 22 torr, respectively.

Although the 1935 expedition put the mine at 5,800 m, a 1972 large-scale contour map of the area (Institute of Military Geography, Santiago) puts it at 5,950 m. We measured the barometric pressure as 373 torr and a reading of 371 torr was made by R.B. Schoene in March 1986. The pressure–altitude relationship is latitude-dependent and the latitude of Aucanquilcha at 21° south is similar to that ot Mount Everest (28° north) where the relationship has been accurately measured[5]. This confirms an altitude of approximately 5,950 m.

Few physiological data are yet available on these men. The barometric pressure of 371–373 torr means that the P_{O_2} of moist inspired gas is only 68 torr. If their alveolar P_{O_2} and P_{CO_2} conform to the pattern previously described for acclimatized lowlanders and high altitude residents[6,7] they would have alveolar P_{O_2} and P_{CO_2} values of approximately 42 and 22 torr respectively (see figure). Arterial blood samples were taken from three of the men in March of this year when they were visiting Ollague (3,700 m P_B 490 torr) and these gave a mean haematocrit of 60.89 per cent which is similar to that of acclimatized lowlanders at a similar altitude[3].

Few people live permanently above 5,000 m even in the Peruvian Andes, or the Himalayan ranges of Nepal or Tibet. The caretakers of the Aucanquilcha mine apparently live at a far greater altitude than anyone else in the world.

I am indebted to R. Santolaya for making the arrangements to visit the mine, to R.M. Winslow for measuring the hematocrits and to R.B. Schoene for a measurement of barometric pressure.

JOHN B. WEST

Department of Medicine, University of California, San Diego, La Jolla, California 92093, USA

1. Keys, A. *Sci. Montly* **43**, 280-312 (1936).
2. Pugh, L.G.C.E. *Brit. med. J.* **2**, 621-627 (1962).
3. Ward, M. In *Mountain Medicine* (van Norstrand Reinhold, New York, 1975).
4. Heath, F. & Williams, D.R. in *Man at High Altitude* (Churchill Livingstone, London, 1977).
5. West, J.B., Lahiri, S., Maret, K.H., Peters, R.M. & Pizzo, C.J. *J. appl. Physiol.* **54**, 1188-1194 (1983).
6. Rahn, H. & Otis, A.B. *Am. J. Physiol.* **157**, 445-462 (1949).
7. West, J.B. *et al. J. appl. Physiol.* **55**, 678-687 (1983).

Functions of leaf fall

SIR—Ford has suggested that leaf fall is the mechanism of excretion in vascular plants[1]. But to view the senescing leaf simply as an 'excretophore' may be to limit our vision. Some species may avoid accumulating high concentrations of heavy metals by removing them in senescent leaves[2]. The dead leaves of other species release allelopathic compounds which play an active part in interspecific competition[3]. And the maintenance — indeed the creation — of the plant's habitat, the A and B soil horizons, depends on leaf fall. The degree to which nutrient elements are withdrawn from leaves before they fall varies between ecological communities and correlates with the availability of nutrients in the soil[4]. Thus elimination of waste products of metabolism may not be leaf fall's only adaptive function.

A contrary view[5] is that higher plants have been able to derive a variety of selective advantages from their dead and dying parts, thanks to their 'modular' form of construction[6]. This allows parts of the organism to be shed without prejudicing the survival of the whole. The result is that as in life so in death and fall, leaves have been subject to many selection pressures and many functions are now achieved by the process.

The practical importance of all this is that it implies that leaf senescence is not simply an avoidable evil. The use of 'anti-senescence' genes or chemicals to increase leaf area duration and therefore yield in agricultural crops, may prove to be not quite as straightforward as has sometimes[7] been assumed.

RICHARD C. HARDWICK

Institute of Horticultural Research, Wellesbourne, Warwick CV35 9EF, UK

1. Ford, B. J. *Nature* **323**, 763 (1986).
2. Waughman, G. J. & Bellamy, D. J. *Ann. bot.* **47**, 141–145 (1981).
3. Sydes, C. & Grime, J. P. *J. Ecol.* **69**, 237–248 (1981).
4. Grubb, P. J. *A. Rev. Ecol. Syst.* **8**, 83–107 (1977).
5. Hardwick, R. C. *British Plant Growth Regulator Group Monogr.* **9**, 61–74 (1983).
6. Hardwick, R. C. *Phil. Trans. R. Soc.* **B313**, 161–173 (1986).
7. Abu-Shakra, S. S., Phillips, D. A. & Huffaker, R. C. *Science* **199**, 973–975 (1978).

Undirectional spin

SIR—When I wrote my paper on undirectional spin[1], which was recently referred to in *Nature*[2], I unhappily had not seen a paper[3] on this topic which preceded mine by a few years. This highly illuminating work, which overlaps mine hardly at all, is particularly to be recommended to your readers for its excellent simulations.

HERMANN BONDI

Churchill College, Cambridge CB3 0DS, UK

1. Bondi, H. *Proc. R. Soc.* **A405**, 265 – 274 (1986).
2. Walgate, R. *Nature* **323**, 204 (1986).
3. Kane, T.R. & Levinson, D.A. *Int. J. non-lin. Mech.* **17**, 175 – 186 (1982).

Detecting Earth-like planets

SIR—Angel *et al.*[1], while advocating the use of a 16 m diameter telescope optimized for infrared observations to search for Earth-like planets outside the Solar System, criticize my suggestion[2] of using an interferometric system, operating at optical wavelengths, for the same purpose. Their criticism supposes that the sidelobe levels would be objectionably high because of the optical imperfections in the interferometry elements but here we point out that there is a fundamental error in their analysis.

Sidelobe suppression is an important issue for both concepts. The interferometer that I discussed would consist of a number of telescopes of diameter D, deployed in an array with spacings larger than D, but not so large that a planetary disk would be resolved. The elements would be pointed away from, but still close by, the star in searching for planets, but photons from the star would cause a noise background because of the finite sidelobe response. The star could be used as a phase reference to stabilize the interferometer, thus allowing coherent integration, and the diffraction sidelobes would be minimized by apodizing the aperture illumination of the individual elements (which were 1.5 m telescopes in the example given). Still, there would be noise sidelobes introduced by surface irregularities in the optics and it is control of these sidelobes that was being questioned by Angel *et al.*

To summarize the causes of noise sidelobes, the usual definition of gain will be used; relative gain (sometimes called directivity) is the gain normalized to unity on the optical axis (this is the inverse of the definition given by Angel *et al.*). The gain contribution from the noise sidelobes at an angle (θ) off the optical axis will be denoted by $G_n(\theta)$. Intensity in the image plane will be the Fourier transform of the auto-correlation of the amplitude error in the aperture-plane phase front. The precise form of the sidelobes will depend upon the error distribution, but as a representative case let the errors have a random distribution, with an auto-correlation of the form

$$<y^2(\tau)> = <\delta^2> 1 - e^{-(2\tau/l)} \qquad (1)$$

for separation τ: ι is a correlation length and is the phase error. If the r.m.s. surface error is ε, then $<\delta^2> = (4\pi\varepsilon/\lambda)^2$. It then follows from the work of Ruze[3] that, for a uniformly illuminated aperture of diameter D with $\delta << \lambda$, the relative gain of the noise sidelobes will have the distribution

$$G_n(\theta) > = 8\pi^2(\iota/D)^2(\varepsilon/\lambda)^2 e^{-(1/2\pi\iota\theta/\lambda)^2} \qquad (2)$$

Physically, the gain distribution will be a speckle pattern extending over an angle $\approx (\lambda/l)$ with individual speckles of size $\approx (\lambda/D)$.

In their remark that "The speckles will look like a cluster of barely resolved stars of 21st magnitude, in which is hidden a 28th magnitude planet" Angel *et al.* misunderstand interferometry. The speckle pattern is derived from spatial frequencies less than (λ/D), whilst the interferometer fringes come from spatial frequencies much larger than (λ/D). Hence the interferometer fringe amplitude comes from the planet and not from the speckles, which only contribute the noise that was accounted for in the signal-to-noise calculation. Their assertion is, therefore, incorrect.

Angel *et al.*, in estimating noise sidelobes from their equation (1) for the quantity they call the gain, illuminate the fundamental issue. Equation (2) shows that the distribution of errors must be considered. The total power scattered from the main beam depends only upon ε, and if, for a given ε, l is small, the scattered power is distributed over a large angle, and the close-in sidelobe level will be low. That is why I chose $l = 1$ cm as the correlation length, giving a sidelobe suppression 1.6×10^3 greater than the example given by Angel *et al.*

On the other hand, if the only errors in the phase front were from large spatial frequencies, their objection would be well-taken. A beautifully polished surface, whose only errors in figure were on large scales, would be unsuitable for finding planets since all the noise sidelobes would be clustered close to the main beam. Thus, the example quoted by Angel *et al.*, who used $l = 40$ cm as the correlation length in their calculation, is concentrating on the wrong regime. Note that, in my discussion, a mirror diameter of 1.5 m was chosen for the receiving elements, because of the extensive industrial experience in making high-quality mirrors in this size range. Industrial suppliers, in informal inquiries, have indicated that achieving this precision of the element surfaces is an economic, not a technical, question.

In summary, it is far too early to choose between the optical interferometer and the monolithic infrared telescope for a planetary search, as both the optical and infrared approaches need more extensive study. The challenges are well-defined and there are no fundamental physical laws that forbid their solution.

B.F. BURKE

Department of Physics,
Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA

1. Burke, B.F. *Nature* **322**, 340 – 341 (1986).
2. Angel, J.R.P., Cheng, A.Y.S. & Woolf, N.J. *Nature* **322**, 341 – 343 (1986).
3. Ruze, J. *Proc. Instit. elect. electron. Engng* **54**, 633–640 (1966).

ANGEL *et al.* reply—Burke has defined a mirror surface by equation (1) with $l = 1$ cm, showing that a relatively large r.m.s. error can be combined with very low sidelobes at 1/4 arc second radius. But this bears little relationship to the best optical surfaces currently made. In practice, errors are not confined to small scales of 'orange peel' and 'dog biscuit'. Zonal figuring, in particular, leaves circular waves on the crucial scale of 40 cm (λ/l = 1/4 arc s). For optical detection of a planet at this radius with apodized 1.5 m mirrors, sidelobes of 10–7 peak intensity need surface accuracy of about 1 Å r.m.s. on a scale of 40 cm, regardless of finer imperfections. By comparison, figuring errors on this scale are about 50 Å r.m.s. in the Hubble Space Telescope mirror. We would find the case for optical planetary detection more convincing if suggestions were given for dramatic breakthroughs in figuring, testing and support methods.

Does the use of an interferometer help, given an overwhelming need for high contrast rather than spatial resolution? For overcoming the photon noise of scattered light it offers at best no advantage over a single dish of the same area. The array gain cannot be better than the value of $D^2\theta^2/8\pi\delta^2$ given in our paper, where D is the diameter of the single dish that collects the same light. The character of the background is the same as the speckle-fringe convolution familiar to observers with ground-based optical interferometers[1], except at an appropriately reduced intensity level. As in the case for the single 8 m dish analysed in our paper, by morphology there is no way to distinguish a planet from the much brighter scattered starlight fringes. Only if the interferometer baseline is 40 m or more will a solar-type star at 4 pc be resolved and the scatterd light modulation be reduced. But an interferometer with so large a baseline, if arranged for good ultraviolet plane sampling, loses efficiency. Much of the planetary energy appears in parts of the diffraction pattern of low surface brightness[2], and the signal to noise ratio becomes much less than optimum.

We conclude that neither the optical spectrum nor interferometry are suited to this problem. The infrared telescope with filled aperture that we described has improved contrast and reduced scattered light for a given figuring precision, and can be accurately aligned by correcting for errors sensed at optical wavelengths.

ROGER ANGEL
ANDREW CHENG
NEVILLE WOOLF

Steward Observatory,
The University of Arizona,
Tuscon, Arizona 85721, USA

1. Hege, E.K., Beckers, J.M., Strittmatter, P.A. & McCarthy, D.W. *Appl. Opt.* **24**, 2565 (1985).
2. Meinel, A.B., Meinel, M.P. & Woolf, N.J. *Appl. Opt. opt. Engng* **10**, 150 (1983).

Meaning from numbers

Ian Stewart

The Rise of Statistical Thinking 1820 – 1900. By Theodore M. Porter. *Princeton University Press: 1986. Pp.333. $35, £23.40.*
The History of Statistics: The Measurement of Uncertainty Before 1900. By Stephen M. Stigler. *Harvard University Press: 1986. Pp.410. $25, £21.25.*

THERE are many histories of mathematics. Statistics, a younger science, has been comparatively neglected. But now we have the happy conjunction of not one, but two fascinating books that describe the origins and early development of what is arguably *the* mathematical tool of this century. Both are scholarly yet readable; both cover the same period; and both are published by Ivy League presses. There the resemblance ends, and the books largely complement one another, the main overlap being the underlying historical reality.

The Rise of Statistical Thinking avoids technicalities and concentrates on the flow of ideas between the natural and social sciences. It emphasizes the philosophical issues raised by novel statistical methods, and how they affected the subject's development. If it has a flaw it is that the technicalities have been smoothed away a little too much, so that one forgets that the founders of the subject were grappling with difficult mathematical, as well as social and philosophical issues.

The History of Statistics is stronger on the mathematical developments involved, especially the early history of probability theory, and is organized a little more explicitly around a selection of key figures. Mathematical detail is especially evident in the first half, where Stigler describes the contributions to probability theory of mathematicians such as Legendre, Laplace, Bernoulli and Gauss. This tends to give the impression that mathematicians have no part in human affairs except by virtue of their mathematics, whereas Laplace, for example, took a great interest in the implications of his work.

The most striking feature of the story is that both the "hard" and "soft" sciences played decisive roles in the development of statistics, and that important ideas and methods were transferred between them — in both directions. To illustrate this in microcosm, consider the "error law" or normal distribution. The name "error law" originated in the work of eighteenth-century astronomers and mathematicians, who, in trying to calculate the orbits of celestial bodies, were forced to take account of the effects of observational error. The idea was transferred into social science by Adolphe Quetelet. He was impressed, for example, by Laplace's probabilistic analysis of whether atmospheric pressure was subject to a regular cause of diurnal variation, and applied such analysis to everything he could think of: measurements of the human body, crime, marriage, suicide. By the early 1870s the idea had moved back into the physical sciences, with the development of statistical mechanics. In his book, Porter points out that both James Clerk Maxwell and Ludwig Boltzmann emphasized the success of the error curve in social science to justify importing it into physics. For example, in 1872 Boltzmann wrote:

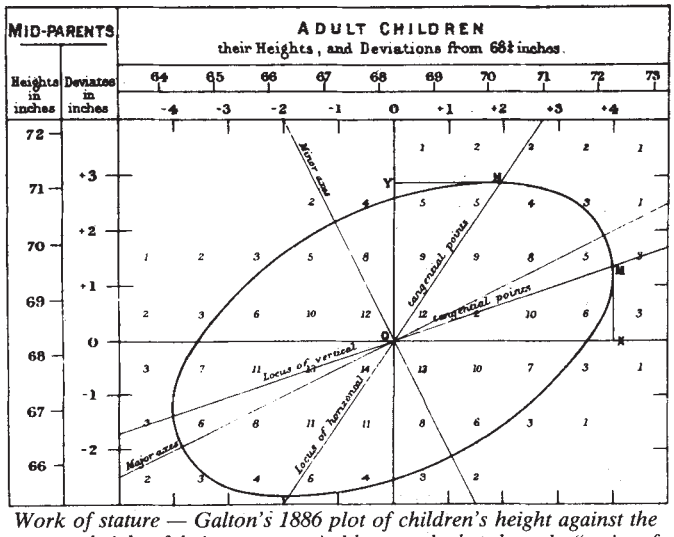
> . . . the molecules of a body are indeed so numerous, and their movements so rapid, that nothing ever becomes perceptible to us except average values. The regularity of these mean values may be compared to the astonishing constancy of the average numbers furnished by statistics . . .

Quetelet was quick to draw general conclusions from the supposed constancy of mean values of social and anthropometric variables, and was the originator of the tantalizing notion of the "average man". He was influenced by the philosophy of Victor Cousin, and the idea of the *juste milieu* which simultaneously emphasized the supremacy of the bourgeoisie and embodied ideals of political moderation. To Quetelet, the "average man" was not just a mathematical abstraction, but a moral ideal. Others disagreed, and the applicability of the error law became highly controversial.

For Francis Galton, like Maxwell and Boltzmann, the error law was a means, not an end. Galton's interest in it arose in a curiously incidental way. A friend of his, William Spottiswoode, wrote a rather dreadful paper applying it to the alignment of mountain ranges. This paper drew Galton's attention to Quetelet's work.

Galton strongly opposed the interpretation of the law that anything exceptional was a flaw. To astronomers, variations in observations were errors, and the aim of the error law was to eliminate them. To Quetelet, deviation from the mean was a sign of moral or physical imperfection. To Galton, however, the variations themselves were the interesting phenomena, and the last thing he wanted to do was eliminate them. Galton, indeed, was the founder of eugenics, and he laid the ground for the "nature–nurture" debate. Always a forceful figure, in 1877 he campaigned to have Section F of the British Association — Statistics and Economic Science — expelled, because it wasn't mathematical enough. He also used the methods of social statisticians to debunk conventional religion: the most notorious of his writings on the subject was *Statistical Inquiries into the Efficacy of Prayer* in which he observed that sovereigns, whose lives were the object of regular prayerful appeal by entire nations, lived no longer than other members of the upper classes. Clergymen had a similar lifespan to physicians or attorneys. Insurance companies charged the same premiums to missionary ships and slave vessels.

Galton's most important contribution, to which Stigler devotes the best part of a chapter, was the idea of regression. Galton could not understand why the degree of variation of a characteristic did not increase from one generation to the next. Surely Quetelet's error law, with its implications of underlying homogeneity, *denied* the possibility of inheritance of characteristics? He eventually realized that a normally distributed mixture of normal distributions would combine to give one overall normal distribution. Each characteristic could follow its own error law. He built a mechanical device, the quincunx, in which lead shot bounced off pins (effectively performing a random walk), to demonstrate the principle. ▶



Work of stature — Galton's 1886 plot of children's height against the average height of their parents, suitably smoothed to show the "series of concentric and similar ellipses" which brought together his ideas about regression in a single elegant mathematical package. (Reproduced from The History of Statistics.)

Comparing the distribution of the heights of children to those of their parents, Galton smoothed the data and obtained a dramatic example of what would now be called a multivariate normal distribution. This led him to the idea of regression lines and gave rise to the analysis of variance.

These ideas were taken up by Francis Ysidro Edgeworth, Karl Pearson and George Udny Yule, who between them took statistics into the twentieth century. According to Stigler, "Even in the varied world of nineteenth-century statisticians . . . Edgeworth was an anomaly". He was a classicist with no higher mathematical training who, while studying law, also undertook a massive programme of self-instruction in mathematics. From the beginning he sought links between mathematical ideas and areas such as psychology, economics and ethics. Soon he began to enquire into the statistical significance of irregularities in observations. At the jubilee meeting of the (Royal) Statistical Society he presented a paper on significance testing for the mean, and went on to develop an even more important idea: the correlation coefficient. Edgeworth changed the focus of statistics. Originally it had been about errors in observations. Now it was about the analysis of structure in data.

The nineteenth-century statisticians placed so much emphasis on the normal distribution that one might be forgiven for assuming that they recognized no other distribution curve. This is not entirely true, but they did expect observations of a single, homogeneous phenomenon to conform to the normal curve. Pearson challenged that viewpoint, and in his first publication on statistics he outlined an extensive theory of skew, or asymmetric, distributions. The full version ran to over a hundred printed pages and led to his election to the Royal Society. Pearson, whom Porter describes as a buccaneering type, was forever trying to grab new territory for statistics; in one memoir he applied his theories of skew variation to 14 examples, including the sizes of crabs, the American divorce rate and the number of petals on buttercups.

Yule was a student of Pearson, and when he arrived at Pearson's laboratory in 1893 he was probably expecting to study applied mathematics. But by 1895, says Stigler, "he was a statistician", and one showing signs of growing independence. Yule applied Pearson's skew curves to data on the frequency of paupers per poor-law union (combination of parishes for administrative purposes). Soon after he applied correlation methods to show that, contrary to assertions of Charles Booth in *The Aged Poor*, there was a relation between (in modern terms) the ratio of direct welfare to work-relief, and the incidence of poverty. In Yule's hands, Galton's eugenic idea of regression became a universal (though still controversial) tool for the analysis of cause and effect.

Thus, by 1900, the main foundation stones of classical probability theory and statistics had been laid. Both authors sensibly stop there, rather than plunging into the technical intricacies of the twentieth century. These are two important and substantial books which, between them, give a great deal of insight into the creation of a fundamental branch of science. In particular it is instructive to witness the way in which general mathematical concepts evolved from a morass of special problems about an enormous variety of topics, at a time when the most difficult question was to find the right questions to ask. □

Ian Stewart is in the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

Northern waters

W.J. Gould & D.G. Masson

The Nordic Seas. Edited by Burton G. Hurdle. *Springer-Verlag: 1986. Pp.777. DM 198.*

"HAFVILLA" is a term used in the Norse sagas to denote getting totally lost at sea. In the scientific sense, *The Nordic Seas* might well provide us with a safeguard against such a fate.

The editor has collected together lengthy contributions from leading authorities in their respective fields, and has produced an impressive reference book on the oceanography (including studies of ice and climate), geology and geophysics of the seas lying between Norway and Greenland.

As is inevitable in such a compilation, there is a degree of duplication of material — for example between the chapters "Brief Overview of the Physical Oceanography" and "The Arctic Waters" — but this is a minor criticism. Over 200 pages are devoted to climatology, ice cover, oceanography (including sound speed) and tides, while the remaining 400 or so deal with geology and geophysics. A surprising omission is any mention of biology, because the subject is of interest to some physical oceanographers and the Nordic Seas are an area in which a great deal of work has been carried out on the relationships between physical and biological indicators and on the effect of physical parameters on fisheries.

The latter part of the volume deals with topography and geology. Much has been written on the geology of the area in recent years but this is the first truly comprehensive review, discussing everything from superficial sediments to sound velocity in the mantle. With its detailed subject index and comprehensive bibliography, it is an excellent "geological dictionary" and will be indispensable for those interested in the Nordic Seas and surrounding areas.

John S. Shelton

Outstanding feature — the 500-m high Ship Rock in New Mexico, a volcanic pipe which was exposed by erosion of its enclosing rock. The picture is taken from M.J. Selby's Earth's Changing Surface: An Introduction to Geomorphology, *published by Clarendon earlier this year. The book will be reviewed in* Nature's *annual textbook issue on 12 March 1987.*

The geological section has its disappointments, however. There appears to be little purpose in the straightforward description of bathymetry (Chaper 9), particularly because it is largely repeated and explained in terms of the underlying geology in the first part of the following chapter. Chapters 10 and 11 are simply too long; much of the considerable volume of data from areas peripheral to or analogous to the Nordic Seas could probably have been omitted without detriment to the overall content. Too much emphasis has also been placed on explanation of even the simplest geological terms — surely inappropriate in a work intended primarily for experts in the field. Finally, although it is otherwise well produced the book suffers from poor reproduction of many line drawings, to the extent that the chapter on bathymetry loses almost all of its impact because contour lines on the charts are invisible and only the annotation remains.

But the merits of the volume greatly outweigh the flaws. *The Nordic Seas* will find its way into the libraries of a wide range of scientific establishments and will be much referred to. It is as up to date as could possibly be expected and will remain as a key document for many years. □

W.J. Gould and D.G. Masson are at the Institute of Oceanographic Sciences, Brook Road, Wormley, Godalming, Surrey GU8 5UB, UK.

Parasitic puzzles

Sydney Cohen

Living Together: The Biology of Animal Parasitism. By William Trager. *Plenum: 1986. Pp.467. $71.40.*

WILLIAM Trager has been a dedicated student of animal parasitism throughout his 50 years on the staff of the Rockefeller Institute in New York. During most of that time the subject was a relatively neglected one. The simpler prokaryote parasites and their host interactions are more amenable to study and received much greater attention. The recent surge of interest in animal parasitism reflects an increased awareness of the global impact of animal parasites as human pathogens, but stems also from several technological advances which have transformed the feasibility of carrying out experiments in the subject. Among such advances is Trager's own most notable scientific achievement, with Jensen in 1976, of cultivating for the first time the human malaria parasite, *Plasmodium falciparum*.

Animal parasites are extraordinarily complex and adaptable organisms. During successive life cycles in different host species, individual parasites may assume forms so disparate as to have been originally regarded as separate species. Alternating life cycles may also involve abrupt changes in metabolism, such as the switch from the glycerophosphate oxidase system of the bloodstream form of African trypanosomes to the cytochrome-mediated respiration of the insect form which is non-infective for its mammalian host. The study of animal parasitism is therefore basically an attempt to understand the sequential control of gene expression which generates alternative parasite forms adapted to recognition, penetration, growth and reproduction in hosts as different as (for example) the sandfly and man. In addition, the student of animal parasitism must attempt to delineate the complex reactions of the host to invasive challenge.

In confronting this daunting task, Trager deals first with a series of fundamental biological events. He starts with parasite recognition and entry into the host, and their attachment to and penetration of specific host tissues and cells. There follows a discussion of the interface between host and parasite and the molecular exchanges across these surfaces that are vital for parasite survival. Host reactions to parasite invasion are considered in relation to innate immunity, the important and often mysterious role of the spleen, acquired immunity and the pathogenesis of parasitic disease. In the latter part of the book, phenomena of parasitism are discussed in relation to a selection of specific protozoan and helminthic parasites of human beings and other vertebrates that are currently the centre of research work. The closing chapters deal briefly with symbiosis and with the methods traditionally used for the control of parasitic diseases, namely, chemotherapy and ecological methods such as vector control.

Throughout the text, Trager draws attention to the innumerable questions raised by the phenomena he describes. In a few instances these can be satisfactorily accounted for in molecular terms. For example, we at last have an answer to the long-standing puzzle of why West Africans and their descendants in the New World are totally refractory to infection with *Plasmodium vivax* and yet are susceptible to the other plasmodial species that infect mankind. Also, the means whereby the sickle-cell gene engenders resistance to infection by malignant tertian malaria is understood in detail. A great deal is also known about the structural and genetic basis of antigenic variation in African trypanosomes which mediates their evasion of potentially lethal immune attack by the mammalian host. For the most part, however, the biological phenomena described by Dr Trager with arresting simplicity provide a Pandora's box of puzzles which will stimulate and challenge future generations of molecular pathologists. The answers — when they come — will illuminate not only the mechanisms of animal parasitism but also our understanding of biological differentiation.

In his introduction to the book, Joshua Lederberg remarks that descriptions of the lives and adaptations of animal parasites are "a veritable Arabian nights of narrative, not of the human imagination, but of Nature's". The subject is indeed a fascinating one, and also is of great consequence in human health and welfare. This book, distilled from a lifetime's work, and set out in uncluttered style with abundant illustrations, will be read with profit and enjoyment by a wide audience including both aspiring and established molecular parasitologists. During the past decade many imaginative initiatives have transformed the study of animal parasitism by bringing together research workers from a variety of disciplines. Dr Trager's book will give further impetus to recruitment into this engrossing and important area of research. □

Sydney Cohen is Emeritus Professor of Chemical Pathology in the United Medical and Dental Schools of Guy's and St Thomas's Hospitals, London SE1 7EH, UK.

Unsure accuracy

Alan J. Charig

The Riddle of the Dinosaur. By John Noble Wilford. *Faber & Faber/Knopf: 1986. Pp.304. £14.95, $22.95.*

ONE might assume from the title that this is just another book on dinosaurs; after all, there are already so many on the market that thinking up a new title must always be a problem. But the exact wording hints strongly at the distinctive nature of this excellent volume. For it is not a descriptive work, listing the various genera of dinosaur and giving an account of the anatomy, physiology and habits of each, nor is it even concerned with their evolutionary history; rather does it deal with the history of dinosaur discovery, with dinosaur research and with the many controversies (both past and present) that bedevil the life of the dinosaurologist.

The text has been well researched and is factually correct. It is also stylishly written, and is *interesting*, even to the professional dinosaur worker. Most important of all, it reports on the controversies in a balanced, impartial manner, often coming to no positive conclusions but leaving the reader sitting on the fence — which, in our present state of knowledge, is in many cases the proper place to be.

The author has based much of his book on interviews and discussions with well-known scientists active in the dinosaur field. Since he himself is an American, a Pulitzer Prize-winning science correspondent for *The New York Times*, it is understandable that those people he chose to talk to were all from the United States; the book therefore gives the impression that dinosaurology today is an almost exclusively American subject. His historical perspective nevertheless embraces a fair amount of material from outside North America; it includes — as it must — the story of the first discovery of dinosaur remains, in England, and the initial perception, by Mantell and subsequently Cuvier, that reptiles of enormous size and (as originally conceived) herbivorous diet had existed in Mesozoic times.

My only complaints are to do with the illustrations, which are extremely few. One of them, the family tree on p. 189, is familiar; I recognize it from another volume, although its source is nowhere acknowledged. In the middle of the book are seven colour restorations by Douglas Henderson; his dinosaurs look wooden and his pictures as a whole have neither scientific nor aesthetic appeal. Moreover three of the seven are given the wrong date.

That notwithstanding, here is a book that no dinosaur addict — be he a professional or one of those innumerable keen amateurs — can afford to be without. □

Alan J. Charig is in the Department of Palaeontology, British Museum (Natural History), Cromwell Road, London SW7 5BD, UK.

Is the reductionist beyond belief?

William Paton

God and the New Biology. By Arthur Peacocke. *Dent: 1986. Pp.198. £10.95. To be published in the United States by Harper & Row.*

"THE brain secretes thought as the liver secretes bile; and you are all just bundles of conditioned reflexes". With this ending to his lectures on the alimentary tract, my physiology professor, the late John Mellanby, leant on his elbow on the lectern, and beamed at us. Although he seemed a pretty implausible reflex-bundle himself, that, 50 years ago, was my first introduction to reductionism. A.J. Ayer's *Language, Truth and Logic* was also giving logical positivism a brisk run. An advanced cleric of the day, L.W. Grensted, was said to believe that "There is no God, men are monkeys and there may be spooks". To the eye today, it was all good clean fun.

Faith, reductionism, the "enfant terrible" and the "advanced cleric" are still with us — but the antitheses are less sharp. The church has learnt to face the criticisms of rigorous scholarship, and science that the ethical problems it can create need ethical principle not of its making for their resolution. Leaders of science and religion are now less prone to pronounce at each other's expense.

Dr Peacocke's book on biological reductionism is admirably consonant with this atmosphere. He is well-qualified; a physical biochemist and tutor in Oxford, who took orders, became a college chaplain in Cambridge and is now director of the Ian Ramsey Centre in Oxford, an institution recently founded for the study of ethical problems in science and medicine. The book pulls together his thinking over a number of years.

His conclusion is best given in his own words (though they are more precise than colourful):

> It *is* possible for higher level concepts and theories... to be non-reducible to lower level concepts and theories, that is, they can be autonomous. At the same time one has to recognise the applicability of the lower level concepts and theories (for example, those of physics and chemistry) to the component units of more complex entities and their validity when referred to that lower level. That is, with reference to biology, it is possible to be anti-reductionist without being a vitalist. Higher level concepts and theories often refer to genuine aspects of reality at their own level of operation and we have to eschew any assumptions that only the so-called fundamental particles of modern physics are "really real". To do so is indeed a kind of "fundamentalism" of a pseudo-scientific kind that is as much to be avoided in impartial enquiry as its religious namesake.

The book is a critical review of a rather extensive literature on and around the "reducing" implications of modern biology, molecular biology, genetics, evolution and sociobiology (but, deliberately, not neuroscience).

Dr Peacocke finds a genuine middle ground. The reductionist refuses to accept any "extra" (for example, entelechy or vital principle) beyond the laws of natural science in order to explain natural phenomena. The anti-reductionist finds it inconceivable that the full variety of biology and human life can be so explained without it. But molecular biology and genetics do bring one striking gift: the breadth of their explanatory power shows how, from simple elements, systems of astonishing complexity and richness can arise. Randomness presents itself not as some irrational intrusion, but as a source of creativity. And it seems that in these new systems, there is a kind of autonomy. Social behaviour, for instance, may indeed depend on the existence of intermolecular forces, but it is to be discussed and experimented on in terms of its own concepts rather than, say, that of the Van der Waals equation. So, too, the B Minor Mass is a "mere" assembly of diatonic tones — yet we have to understand and discuss it in musical, not acoustic language. If, still on the threshold of knowledge, we can already see that in the primary chemical elements there lies the potential for the synthesis of proteins and enzymes, and cell replication, cell-constituent assembly, speciation, learning and association, what may there not be still to be revealed?

Like other intermediate positions, it may not satisfy. Does it leave, for the Christian, room enough for a God with a loving personal relation to his creation? Does it provide, for the reductionist, too much room for self-deceiving religious aspirations to creep in? Perhaps it is not so much the possible existence and human awareness of some higher being that can most worry a scientist, but the idea of that being interfering with the natural course of events, with "natural law". Yet the question is not much different from the age-old one of whether our own day-to-day decisions (seemingly "free") actually modify events or not. If our belief in the one influence is valid and compatible with natural law, why not the other?

Dr Peacocke, though unequivocally Christian, is not, I think, pressing his views. His posture is like that of Bishop Butler in the *Analogy of Religion*, published almost exactly 250 years ago:

> It is come, I know not how, to be taken for granted, by many Persons, that Christianity is not so much as a Subject of Enquiry; but that it is, now at length, discovered to be fictitious. And accordingly they treat it, as if, in the present Age, this were an agreed Point, among all People of Discernment; and nothing remained, but to set it up as a principal Subject of Mirth and Ridicule, as it were, by way of Reprisals, for having so long interrupted the Pleasures of the World. On the contrary, thus much, at least, will be here found, not taken for granted, but proved, that any reasonable Man, who will thoroughly consider the Matter, may be as much assured, as he is of his own Being, that, however, it is not so clear a Case, that there is nothing in it.

For those who do not like to see people bludgeoned into belief, and who are also interested in the work of recent years by both a theologian and a scientist, Dr Peacocke's eirenic writing will be much welcomed. □

Sir William Paton, 13 Staverton Road, Oxford OX2 6XH, UK, is Emeritus Professor of Pharmacology at the University of Oxford.

THE MANAGEMENT OF AIDS PATIENTS

Edited by DAVID MILLER, JONATHAN WEBER and JOHN GREEN

The Management of AIDS Patients is the first comprehensive guide to the practical clinical management of patients with AIDS or HTLV III infections. The book avoids the sensational aspects of the disease, offering solid advice and information for all people involved in patient care.

The editors and many of the contributors come from St Mary's Hospital, London one of the foremost centres in Britain for the treatment of AIDS patients. The knowledge and experience of these experts has been combined to provide a much-needed book for doctors, nurses, dentists and indeed for health-care professionals everywhere.

1986

200pp	Hardback	0 333 40465 3	£30.00
	Paperback	0 333 40466 1	£10.95

Please order this title from your bookseller – or in case of difficulty from Judy Chappell, Macmillan Press, Houndmills, Basingstoke, Hants RG21 2XS.

MACMILLAN PRESS

New in paperback

- *The Birth of Particle Physics* edited by Laurie M. Brown and Lillian Hoddeson (reviewed in *Nature* **308,** 383; 1984). Publisher is Cambridge University Press, price is £12.95, $18.95.
- *The Ecological Web: More on the Distribution and Abundance of Animals* by H.G. Andrewartha and L.C. Birch (reviewed in *Nature* **315,** 81; 1985). Publisher is University of Chicago Press, price is $19.95, £16.95.
- *Measuring the Universe: Cosmic Dimensions from Aristarchus to Halley* by Albert Van Helden (reviewed in *Nature* **316,** 492; 1985). Publisher is University of Chicago Press, price is $8.95, £7.50.

The music of digital computers

Philip Campbell

More and more composers are relishing the unprecendented freedom provided by digital processors to manipulate both internal and external characteristics of sounds.

AMONG the most laughable events in the history of music—according to the cartoonists of the time at least (Fig. 1)— were the concerts of the Italian futurists during the first decades of this century. Depicted thus as sonic anarchists, contemporary photographs and manifestos nevertheless indicate the high seriousness with which, with the help of their 'noise-intoners' ('*intonarumori*'), the futurists fused the sounds of everyday machinery and mechanically synthesized noise into no-doubt outrageous cacophony. Though that artistic movement proved still-born, more recent figures, driven by a similar determination to escape what they see as the shackles of the past, have provided the composer with an instrument of more profound potential—the digital computer, capable of dissecting, transforming or synthesizing sounds according to a limitless variety of prescriptions.

The association of the computer with music goes back several decades. Composers such as Milton Babbit have used it since the 1940s to create ultra-rigorous twelve-tone music while, at another extreme, Iannis Xenakis has used it to inject a more statistical component into musical design[1]. Others have used computers for musical analysis and for modelling aspects of musical perception (see, for example, ref. 2).

But it is only now that the capacity of the digital computer is beginning to approach that required to make it a truly flexible tool for the creative musician—not a crutch by which deficient composers can save themselves work but a doorway to new paths of musical imagination, in particular those that encompass the microstructures of sounds.

In this article I shall concentrate on two particularly stimulating and poetic realisations of the computer's potential. Both, as it happens, were produced at what is currently the mecca of computer-assisted music, the Institut de Recherche et de Coordination Acoustic/Musique (IRCAM)—a subterranean establishment adjacent to the Pompidou Centre in Paris. The first, *Mortuos Plango, Vivos Voco* by the English composer Jonathan Harvey, is a composition on tape, about ten minutes long, which manipulates and intertwines only two sounds—a bell and a singing voice. The second, *Répons* by Pierre Boulez, is yet to be completed but, through several preliminary exposures, has already established itself as a benchmark in the use of computers to transform the sounds of musical instruments during a live performance.

Consider the scale of the problem. Beethoven's ninth symphony reaches our ears in the form of a complex pressure signal lasting about one hour. The frequency range of the best human ears extends upwards of 16 kHz; to satisfy them, a digital device wishing to reproduce the aural magic of, say, the Berlin Philharmonic must be able to synthesize or reproduce waveforms up to such frequencies—we can safely set the upper limit at 20 kHz. The principle of digital operation involves breaking up the continuous signal into a series of discrete samples and, because one has to have at least two samples per oscillation to reproduce a wave at a given frequency, we require a sampling rate of 40 kHz. Furthermore, the amplitude of the waveform as it varies from sample to sample must be resolved with good precision, otherwise noise ('quantization errors') will be introduced. Sixteen bits are usually considered as an adequate representation for each sample ($2^{16} = 65536$).

Thus simply to store Beethoven's ninth symphony in a digital memory requires about 14×10^9 bits; to synthesize it from scratch would represent a truly gargantuan feat of calculation (and perhaps a sterile one at that). For any instrument one might require a digital oscillator for each harmonic component of the waveform, each of a relative amplitude that, as it happens, varies according to the pitch and loudness of the note being played. Furthermore, to sound natural each harmonic component has to have the appropriate relative phase together with a small degree of random variability. Finally, the resulting tone has its own 'global' envelope determined by the composer's score, variable loudness as in a crescendo, for example, or perhaps a long sustained note of rapidly re-iterated attacks, a tremolando.

It is clear that to provide a flexible digital instrument for the composer requires truly formidable computer power. Necessarily, both of the pieces to be discussed here make do with considerably

Fig. 1 A futuristic evening (Boccioni, 1911).

less than the full real-time synthesis described above. The tape of *Mortuos* was painstakingly accumulated over months of work in a studio (there is no musical score—to perform the piece one simply plays the tape). The live transformations of *Répons* primarily involve relatively simple arithmetic operations on incoming signals. Of course the musical power of such techniques springs not from their degree of sophistication but from the sureness of judgement with which they are married to strong ideas.

The music of *Mortuos*

In the compact Norman tower of Winchester Cathedral can be found a set of seventeenth century bells, on the largest of which—the tenor C bell—are inscribed the words "*HORAS AVOLANTES NUMERO, MORTUOS PLANGO: VIVOS AD PRECES VOCO*" ("I enumerate the fleeing hours, I mourn the dead, I call the living to prayer"). Two sounds provide the basic elements from which *Mortuos* is constructed: the peal of this bell and the treble voice of the composer's son singing these words.

The publishers of *Nature* are, alas, insufficiently generous to provide the reader with a complimentary recording of *Mortuos*. Moreover the only commercial recording currently available[3] is unlikely to be found on the shelves of any but the most enlightened record stores (of which there appear to be none in London, for example). In the face of such obstacles, the following description of the music will have to suffice.

Any attentive reader of newspapers will have noticed the old trick by which his or her attention is sought—the classic first paragraph in which the essence of a story is presented with maximum impact. Harvey achieves precisely this, opening *Mortuos* with a vigorous carillon of bells which gradually declines in intensity until only two or three slowly pealing bells are left, over which the boy intones part of the inscription from which the title is drawn. Both bell and boy sounds, here and throughout the piece, are digitally synthesized or reproduced.

The significance of this carillon is greater than might be thought at a first hearing, when it is likely to appear as a random clustering of bell sounds. As it happens the composer is here representing in macroscopic terms the microstructure of the tones of the C bell. The harmonics produced following a bell stroke, which Harvey analysed (4) by a standard fast Fourier Transform, are shown in Fig. 2. During the reverberation these harmonics decay in descending order. This microprocess is writ large in the opening carillon in that each of the bells that simultaneously sound is repeatedly playing one of these harmonics; the higher the pitch of the bell, the more rapidly it repeats its stroke and the sooner it ceases, giving rise to the cumulative and compelling effect described.

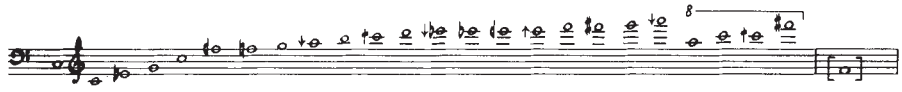


Fig. 2 Spectrum of the tenor C bell of Winchester cathedral in musical notation. The final F in parentheses is a fictitious harmonic, apparent only to the human ear deceived by upper harmonics[4]. Key: ↗ sharpened (raised) by a quartertone: ♯ sharpened by a semitone; ♩ flattened (lowered) by a quartertone; ♭ flattened by a semitone; ↑ adjusted by less than a quartertone.

Each of these 'bells' was synthesized with a spectrum identical, in relative terms, to that in Fig. 2, being simply a reproduction of the C bell transposed upwards in pitch. The reverberating tones of the C bell were digitally sampled and stored in a 'wave-table' within the computer memory which could itself be sampled at varying speeds, both backwards and forwards. This facility allowed Harvey not only to generate the opening carillon but also, later in the piece, to manipulate and overlay the pattern of decaying harmonics to yield subtle rhythmic effects or, by simply reversing the order of fading, to 'turn the bell inside out'[4].

Having caught his listener's attention with this declamatory exposure of his basic materials, Harvey proceeds to unfold the only element of the composition which might be called a 'theme': the slow succession of notes drawn from the C bell spectrum, shown in Fig. 3. Each note is played by bell tones 'faded' in and out so that successive notes overlap, their combined spectra providing a high frequency sheen above the notes. Such effects of timbre, together with harmonic juxtapositions such as that between an A♭ in the spectrum of the note F and the following A♮ of the "melody", all in the context of the quiet simplicity of the theme itself, combine to yield a memorable and beautiful statement.

This gesture opens the first of the eight principal sections of the piece, each of which begins with one of these thematic notes struck by a bell: within each section that note provides a basis for computer-aided musical development. Thus the entire piece adds up to an elaboration of the theme, all built with the materials presented at the outset.

In one sense the music is tonal (in contrast to the atonal style prevalent in music written over the last two or three decades), in that a pitch from which the material springs is fairly explicit at any point in the composition. The intervals within the primary bell spectrum, particularly the fourth, fifth and octave, pervade the piece—though a unique and essential flavour is added by those notes most removed from the well-tempered scale. But at no point is the writing harmonic in a progressive fashion (as in the classical style, where one chord leads to another via movement of the bass). In that sense the piece is static—a contemplation, if you like, of the inner life of certain sounds. Nevertheless, the flow of Harvey's invention creates a momentum of its own.

The techniques by which each of the thematic notes is elaborated within each section represent something of a showcase for the potential of the computer for sound sculpture. For instance the second section (based on G) leads to the third (F) by an unprecedentedly literal means of harmonic preparation: from a chord comprised of a combination of boy and bell tones, some of the notes suddenly slide upwards and/or downwards in pitch (an effect known as a glissando) to arrive at a new set of notes forming part of the spectrum of the F bell stroke which follows. Here it is the synthesized voices of the boy which execute the glissandi whereas, in the third section, it is the sine wave components of the bell that do the work in what turns out to be a study of glissando manipulation.

In the fifth section such simultaneous glissandi, executed in clusters of voices with added reverberation, provide a spectacular quasi-choral backcloth. Harvey organises and overlays the glissandi so as to generate a sense of chordal progression behind fragments of text sung solo in the foreground.

Each section of *Mortuos* exhibits at least one distinct facet of the digital computer's ability to render malleable the internal as well as the external characteristics of sounds, varying the juxtaposition of the harmonic components such that bell and boy at times merge imperceptibly into one another—with the bonus of spatial effects added by means of the eight tracks of the tape which can be independently routed to different speakers.

The eighth section provides a final example. Here the synthesized C bell peals with slow repetition while a choir of boys

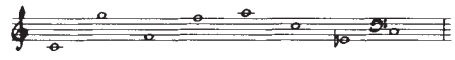


Fig. 3 The eight primary pitches from the bell spectrum which define both the 'theme' and the principal sections of *Mortuos*.

periodically 'tolls' a dissonant note cluster derived from the C bell's spectrum. But the harmonics of the bell are differentiated by the manner in which their loudness is varied. Some are sustained over many strokes while the majority follow the pattern of impact and subsequent decay, but with added slow variation within the overall decay envelope. This simulates the effect of the swing of a bell—for a static listener, the perceived spectrum varies cyclically with the changing orientation. And in an inspired final gesture, Harvey removes the initial sounds of the impacts, maintaining the 'swinging' effect within the fading peals. It is as if the air itself were echoing the vanished bell.

Synthesis in *Mortuos*

The techniques used by Harvey (together with his technical assistant at IRCAM, Stanley Haynes) represent an appreciable proportion of those currently available for digital synthesis. (For an excellent overview see ref. 5 or refs. 6 and 7). Once the harmonic components of the C bell had been determined, its sound could be synthesized by combining sine tones of appropriate frequencies and relative amplitudes and phases—a technique known in the trade as additive synthesis. Another approach adopted, known as fixed waveform synthesis, was to store the digitally sampled sound in successive memory locations—the wave table technique already mentioned. Finally subtractive synthesis, in which a broad-frequency-band source is subjected to appropriate filtering, was used to reproduce the boy's voice.

Additive synthesis. The complexities of building up the highly varied sounds of musical instruments from their components have already been mentioned. Bell sounds, at least in their natural form, are somewhat simpler to reproduce. By means of the Music V software at IRCAM[8,9], Harvey was able to generate both the natural bell sounds and the control of individual harmonics described above. Additive synthesis also permitted interpolation between two tone qualities or timbres at a given pitch; thus by appropriate variation of the harmonics a bell sound could gradually be transformed into a treble voice or vice-versa. Furthermore, by sending different harmonics to different speakers the composer was able to conjure up the illusion of being inside the bell.

Fixed waveform synthesis. By storing sound in a wavetable, one loses the ability to manipulate its harmonic components but gains significantly in the ease with which the sound as a whole can be transformed. For example one can easily transpose the sound to a higher pitch by increasing the effective rate at which the waveform is sampled. For a given rate of access to the memory, one can double the frequency (transpose up an octave) by sampling every other location along the wavetable. Of course this procedure distorts the waveform, making its envelope less smooth and thereby introducing noise, but sampling rates are typically high enough for this not to matter. Another problem occurs if there are harmonic components present whose frequency is higher than half the effective sampling frequency. Such components will be insufficiently sampled and will lead to "aliasing"—the generation of spurious signals at lower frequencies.

Subtractive synthesis. One of the long-standing goals of sound synthesis is that of mimicking the human voice—the ultimate user-friendly computer will of course use speech synthesis at a level indistinguishable from that of humans. The approach adopted by most working in this area is that which our species has evolved: a sound source subsequently modified by a series of resonators that provide individual characteristics as well as common elements of language such as vowel sounds. These manifest themselves as distinctive features in the spectrum superimposed on the fundamental tone of the speaking or singing voice. The source in the human case is a double reed (the vocal chords) which modulates a stream of air travelling from the lungs through the trachea, a tube of continuously varying diameter which in turn feeds several cavities including the mouth and nose. The lips, tongue, palate and glottis can introduce turbulence or sudden perturbations into the flow to produce 'fricative' and 'plosive' sounds respectively. The aim of subtractive synthesis is to mimic this system with appropriate sound sources and filters.

To reproduce his son's voice for *Mortuos*, Harvey used the CHANT algorithm developed at IRCAM by Xavier Rodet and others[10]. This digital approach generates fricative and plosive sounds by means of a flat-spectrum source coupled to appropriate filters, generating a series of pulses spaced according to the desired fundamental frequency. This output is then fed through a more complex set of filters to reproduce the effects of the resonant trachea, larynx, mouth and other cavities. (At the time that *Mortuos* was written the CHANT algorithm could not simulate consonants, which were therefore digitized directly from the boy's voice).

The design of filters, whether analog or digital, is a vast subject in itself. Just as in a simple electronic circuit it is far from intuitive to decide what arrangement of capacitance, inductance and resistance will provide a specified frequency response, so the design of filters can require detailed calculation straying into the complex plane of imaginary numbers, for example.

Digital filters need not be complicated in practice, however. By simply taking a running average of pairs of successive samples, for instance, one smooths any sudden change in amplitude thereby effectively applying a low pass filter. More generally, by adding or subtracting appropriately weighted input and (previous) output samples one can apply a filter of almost any desired characteristics. Finding the right arithmetic combination is the problem. An alternative approach, adopted in the CHANT program[10], is to use each pulse of the signal from the sound source to generate, in parallel, the calculated impulse responses of the desired filters. These responses, which take the form of exponentially damped reverberations decaying over several sample periods, can simply be read from memory and appropriately summed for each output sample.

Perhaps the most striking feature in *Mortuos* is the apparent ease with which, with the help of these digital techniques, two totally different sources of sound (one inflexible and metallic, the other variable and human) are bonded together as complementary acoustic entities, exhibiting or merging their individual characteristics at the composer's will.

Computation in performance

Répons, the latest composition of Pierre Boulez, is something of a landmark, not least for the degree of technical and musical integration between computer and musicians. As is typical of a Boulez composition, the piece has driven music critics to extremes: ". . . spasmodic gasps and lurches . . . no purpose . . . at odd moments it was superficially stimulating as a cold shower can be. . ." (*New York Times*); ". . . a great blinding sun of a piece, commanding its satellites in majestic orbit . . . the supreme exercise of the intellect becomes the supreme gratification of the senses . . . reverberating like bell towers of a mighty celebration" (*Boston Globe*). As it happens, no-one (other than the composer, perhaps) has heard the work in its entirety. The musical score is complete, however, and many of the computer transformations have been written, and have been heard in a number of preliminary performances.

The title of the piece, derived from the words *réponse* (reply) and *reponsoire* (responsory), reflects the composer's focus on relationships between a soloist and a group. This idea applies both literally to the ensemble of performers—six soloists responding to a 24-strong group of instrumentalists, one computer responding to the soloists—but also figuratively to the music itself. The aural clarity of much of the writing and the sheer drama of many of the computer transformations combine to yield a formidable impact in the concert hall.

The physical layout of a performance

of *Répons* is itself unusual. Six soloists (two pianos, harp, vibraphone, glockenspiel and cymbalom) and six loudspeakers are distributed around the performing space while the intrumental ensemble (woodwind, brass and strings), electronic processors and conductor are placed centrally. Only the soloists, connected to the processors by acoustic and contact microphones, are electronically transformed although their natural sounds are, of course, also heard.

The soloists have certain sound characteristics in common. Each note, being produced by struck or plucked strings or a metal bar, consists of a sudden attack and resonant decay, such that every instrument can be subjected to the same type of computer transformation. The 13,000 or so lines of computer code already used in the piece produce a variety of transformations such as modulation of one instrumental tone by another, controlled retardations to produce rhythmic patterns, frequency shifting and selective amplification and mobile distribution of sounds around the speaker system. (Boulez specifies very precisely the effects that he wants. His instructions are then expressed in computer code by his assistant Andrew Gerzso.)

The technology of *Répons*

The signals from the six soloists feed via a standard mixer into a 'matrix'. This 16 × 16 network can be programmed to adopt particular configurations by which, for example, one of the input channels can be fed into several outputs or vice-versa; furthermore these analog channels can each be amplified or attenuated. The matrix, controlled by a 68000 VME computer, distributes processed signals around the speaker system, again according to preprogrammed patterns which are activated as the performance proceeds. The matrix is also responsible for feeding signals to and from the analog-to-digital (AD) and digital-to-analog (DA) converters which form the interface with the core of the system, the 4X sound processor which is itself also controlled by a 68000 computer.

Designed by ex-nuclear physicist Giuseppe di Giugno, the 4X is the most powerful sound processor currently available, the speed and scope of which stem from its eight separately programmable processor boards[11]. Each of these consists of a unit for arithmetic and logical operations (the latter manipulating the 24 individual bits that make up each binary word, the former using those words as standard binary numbers), a multiplier and three main memories: 64,000 words, each of 16 bits, for the wave-table, 1,024 words of 48 bits for program storage and 1,000 words of 24 bits for program variables. Also within the 4X there is a control board, including 256 programmable timers which can, for example, initiate exchanges of program segments or memory with the 68000 host computer at prearranged times.

For all its sophistication, the 4X has a number of shortcomings which the next generation of machine—the 5A, scheduled for completion in 1989—should overcome. As well as being physically more compact and portable as a result of its use of very large scale integrated circuits, the 5A will have better data width or resolution (32 as opposed to 16 bits), a higher sampling rate (40 rather than 16 kHz) and larger storage capacity.

A more subtle improvement will follow from the use of floating point rather than integer arithmetic. Calculations involving floating point representation are slower but the improvements in signal-to-noise ratio arising from increasing resolution (or decreasing quantization error) are more substantial. The extra data width will provide another benefit: although the resolution of the AD/DA converters may be 16 bits (which is entirely adequate for most purposes) a higher resolution is needed within the calculating units to minimise the effects of rounding errors which accumulate over successive arithmetic operations, potentially disastrously if only 16 bits are used.

For any particular section of the musical score of *Répons* there is a corresponding segment of computer program, keyed in at the appropriate moment of the performance by typing in a single stroke at the keyboard of the 68000 VME. Throughout that section of the piece, while the players and conductor perform the musical score in the usual fashion, any signal from a particular soloist will be transformed automatically by the respective 4X processing board, subject to independent routing or attenuation by the matrix. Boulez intends to develop more sophisticated and multipath programs for subsequent compositions.

The music of *Répons*

Of all the computer-assisted musical gestures used by Boulez, by far the most arresting and theatrical occurs at the point where the soloists and the 4X first enter the proceedings (score number 21 in ref. 12). This entry follows an extended introductory exposition in which Boulez presents musical metaphors for the title of the piece, including most noticeably sections where certain pitches attain particular prominence while chords and textures shift about them—a procedure equivalent to the 'pedal' in classical harmony. The subsequent music sets the instrumental ensemble and the soloists in a continual dialogue, the two groups performing contrasting musical materials.

This pedal device again assumes prominence at the soloists' entry, represented by a massive chord played by all six in arpeggiated fashion (that is, the notes contributed by each soloist are struck or plucked in rapid ascending order). Different soloists are amplified through different speaker combinations adding to the spatial impact of the gesture. As this chord decays, successive soloists each play an answering chord, the respective chords having top and bottom notes in common (see Fig. 4), adding a degree of aural focus to the electronic manipulation which is also applied at this point.

Two computer transformations operate here: time delays and frequency shifts. The delays are applied simply by storing every digital sample into a wavetable as it arrives and then retarding the "pointer" by an amount corresponding to the milliseconds of delay required. Each delayed signal can be individually attenuated.

A selection of these delayed signals are fed through a variety of digital frequency shifters. By appropriate manipulation of a group of successive samples these add or subtract a specified frequency to every harmonic component of the signal[9].

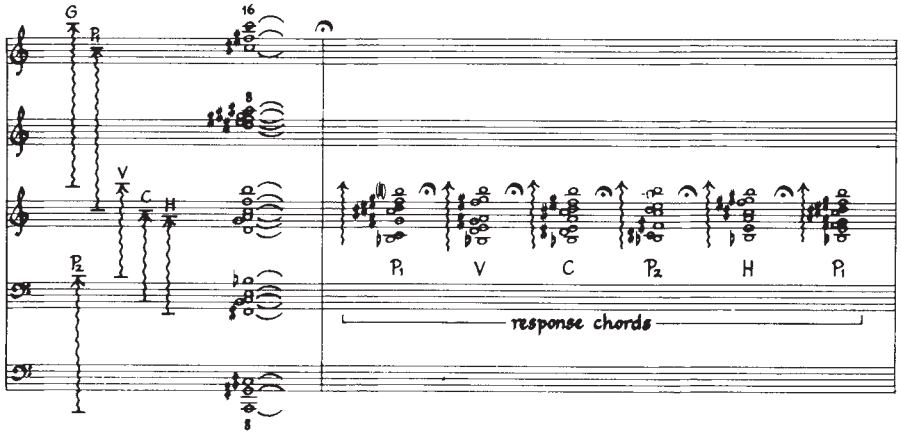


Fig. 4 Entry of soloists in *Répons*, reduced score. All chords are played as loudly as possible, each separated by several seconds. The arpeggiated contributions of each soloist to the initial chord are indicated: P1,2, pianos 1 and 2; H, harp; V, vibraphone; C, cymbalom; G, glockenspiel. Note that the part of the initial chord played on the vibraphone is of particular significance (see text). The six chords that follow are subjected to extended computer transformations.

Because the musical scale is logarithmic (for example, every octave step upwards in pitch represents a doubling of frequency), the linear transformation of frequency addition or subtraction will completely alter the harmonic relationships and hence the character or timbre of the sound. In many cases the result is reminiscent of tuned metal percussion instruments from the Far East—the Balinese *Bonang* or the Chinese *Yün-lo*. Despite the diminished "purity" of the spectrum, the fundamental pitch is usually clearly audible.

The manipulation of pitches, particularly within chords, is fundamental to the musical and electronic writing in *Répons*. One chord in particular appears to contain the roots of much which follows: that played by the vibraphone within the first entry chord (see Fig. 4). This appears in a prominent form at the very opening of the piece; moreover the surrounding components of the entry-chord, each of the answering chords played by the soloists and some of the frequency shifts are derived directly from it[13]. This 'primary' chord also appears prominently elsewhere.

An attempt to represent the combined effect of delays and frequency shifts, as specified by the *Répons* software for one of the soloist's response chords, is shown in Fig. 5.

These techniques are applied at several other points in the composition but the aural images that result vary considerably according to the musical context. For instance, at one point all the soloists are gathered together (in the bar preceding number 27 in ref. 12) to play the primary chord discussed above. This recapitulation of the chord is of some significance within the overall form of the piece and is certainly exhibited as such: it suddenly emerges at full volume from the ensemble of soloists, being sustained by tremolando with gradually decreasing loudness over several seconds—a significant duration in this context. The solo tremolandos are all fed through the same sound processor with several delays being applied, each delay with its respective frequency shift and each directed through a different speaker. Thus the scope of this already powerful moment is enhanced in pitch and space.

One technique not mentioned so far, and yet one which dominated earlier developments in analog music studios, is ring modulation. Here two signals are simply multiplied together. For two sine waves this results in the suppression of the two waves and the generation of two product waves or sidebands whose frequencies are the sum and the difference of the originals. For any instrumental sound one effectively multiplies every harmonic component by a modulating sine wave of specified frequency, which can result in very "thick" tones, packed with partials (spectral components which are not harmonically related to the fundamental) and bearing no resemblance to the original tone. The principal interest of ring modulation arises from the subtlety in the choice of the modulating frequency. If this frequency is harmonically linked to those of the instrumental tone, the resulting partials overlap to yield a 'cleaner' spectrum.

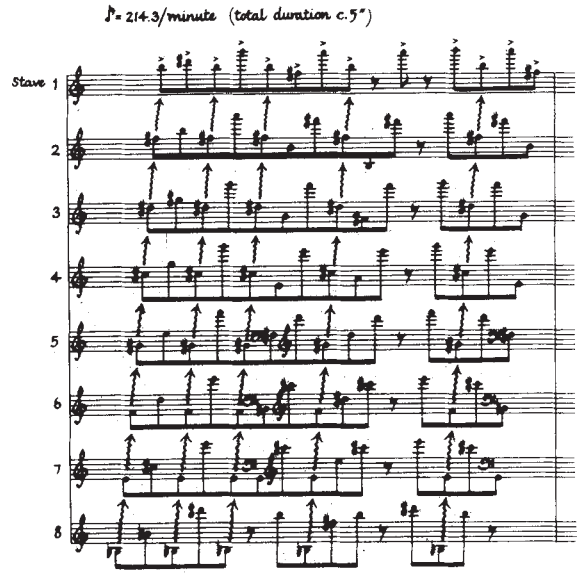


Fig. 5 Computer transformation generated following the final chord shown in Fig. 4. Each stave (from 1 to 8) contains the notes (or their nearest semitone approximations) that are generated by each of the eight notes in the chord. Thus the first note of each stave—linked from stave to stave by arrows—forms part of the arpeggiated chord played on the piano. Subsequent notes linked in similar fashion form the same chord delayed without frequency shift. All other notes are the products of delays and frequency shifts. The 14 delays applied are regularly spaced at 0.28 seconds. The uppermost initial note is emphasized by the pianist in performance, so that its transformations dominate the texture. The ear does not follow each line but resolves the texture into layers of rhythmic patterns at different registers (bass, treble and so on). This combined with the exotic timbres resulting from the frequency shift gives an overall effect of multiple echoes of the piano's chord in the midst of a rhythmic gamelan-like texture. This diagram was derived from the computer code of *Répons*.

One particularly complex example of this approach occurs in *Répons* (at score number 53, ref. 12). Here Boulez gives each soloist a different collection of seven pitches to be played in a variety of orders and rhythms. In designing the ring modulations for this passage, the composer treated these notes as eighth harmonics of respective fundamentals, choosing as modulating frequencies the nth harmonics where $n = 2$ for the first piano, 3 for the vibraphone and so on. The groups of modulating frequencies are injected on the respective processing boards of the 4X. Simultaneously, Boulez specifies several optional layers of transformations with differing combinations of delay and ring modulating frequencies. The options vary from instrument to instrument, the particular combination of layers being chosen in part by a random number generator, resulting in the most elaborate transformations in the entire piece.

Such complexity, which if excessively applied leads to tired ears however fascinating the detail, is preceded and contrasted by a simpler but still striking device, in which every soloist rapidly reiterates (albeit with ornamentation) a single note (B below middle C—see score number 52, ref. 12), thus providing an excellent opportunity to throw the computer's role into sharp definition. Here it acts as a ring-modulating 'light-house': a series of 'flip-flop' switches within the software scans around the soloists, applying time-varying modulation frequencies giving rise to rapidly varying spatial distributions and bands of transformed tones diverging upwards and downwards in pitch.

The final moments of this extensive piece refer back to the first entrance of the soloists discussed above, both in its musical elements and in the combination of frequency shifts and delays used by the computer. As was shown in Fig. 5, the use of arpeggio in the soloists' chordal responses gave rise to markedly rhythmic sequences of notes overlain out of phase. Here, in contrast, the transformed chords are played without arpeggio. The final gesture of all, for example, comes from the computer as it reflects—via a few delays and frequency shifts—a single quiet piano chord. The lack of arpeggio in the initial attack results in a well-defined (albeit brief) rhythmic series of transformed chords, a surprisingly simple and gentle end to a massive and complex composition, and all the more effective for that.

The past, the future

This essay may be considered by some to be something of an insult to the general discipline of computers and composition—no serious attempt has been made, after all, to set *Mortuos* and *Répons* within their historical and contemporary context. For example, the work of the Italian futurists pales into complete insignificance when compared with that of the equally radical pioneer of *musique concrete*, Edgard Varèse. He, at least, held a vision whose fulfilment may now be seen on the

horizon. "I dream of instruments obedient to my thought" he wrote, "and which, with their contribution of a whole new world of unsuspected sounds will lend themselves to the exigencies of my inner rhythm". And for years such centres as Stanford, MIT, Bell Laboratories and institutes in Paris, Cologne and Milan have developed the hardware, the software and the music on which this fulfilment is being built.

For all that, it seems that such studies have reached a turning point within the general context of contemporary music. A substantial proportion of the most gifted young composers of the 1950s and 1960s were obsessed with musical language (and necessarily so), to the extent that uncompromising explorations alienated the bulk of the public audience traditionally associated with 'serious' or 'art' music. Only a few, the most notable being Karlheinz Stockhausen, found the world of electronics appropriate to their modes of communication (or, in too many cases, the lack of it). Thus the combinations of exploratory languages and techniques pushed composition with electronics into a ghetto within a ghetto.

In contrast, the powerful techniques becoming available for digital synthesis, and compositional achievements such as those of Harvey and Boulez at IRCAM, have recently stimulated a degree of interest among the wider community of composers that is unprecendented. Given the technical facilities and assistance that IRCAM is now providing to composers new to electronic techniques, and given that many of these composers are writing music that is less forbidding than much written only fifteen years ago, the time must rapidly be approaching when the widest musical public will be able to encounter the best of this work with ease.

This enhanced degree of interest from composers is likely to throw into relief all the more vividly the questions—both practical and theoretical—that are raised by the use of computers. For instance, what is their potential formal significance in live transformation? The devices used in *Répons*, for example, are purely passive in formal terms in that they produce reflections of composed material. These aural images add an indispensible component to the fabric of the piece but not to its essential form—they are the equivalent of a cathedral's windows rather than its weight-bearing walls and arches.

But the potential is there. As a move in that direction, Boulez intends to return to a previous composition, Explosante-Fix, which incorporated (unsatisfactorily in his view) live electronic manipulation. The piece will be revised so as to include reactive manipulation of the structures of chords as the performance progresses.

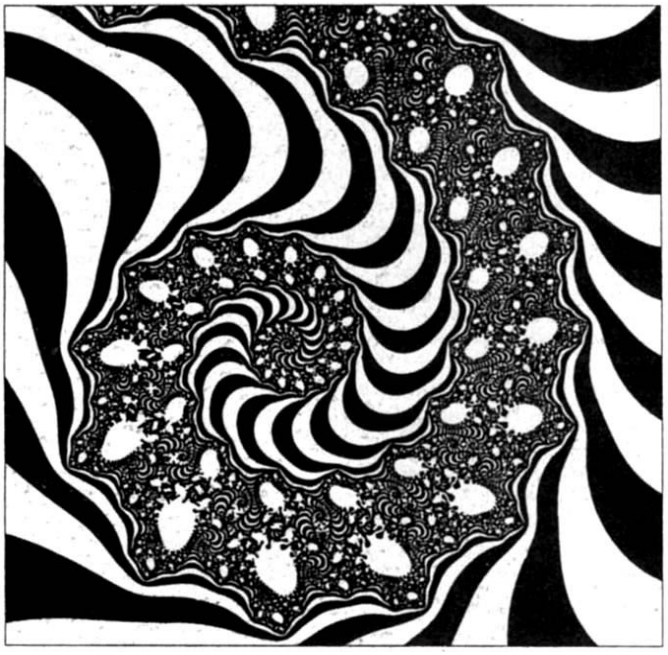


Fig. 6 A highly expanded view showing the border of the Mandelbrot set for the operation $x_{n+1} = f(x_n) = x_n^2 + c$, as mapped onto the complex c plane. The black and white areas demarcate values of c that lead the process $f(x)$ to different regions ('attractors') of the complex x plane[14].

And there are more extreme possibilities. One can imagine an abstract composition whose micro- and macro-structures derive entirely from mathematical formulae exploiting relationships between pitches, harmonics and rhythms for example. Before dismissing such a scheme as depressingly arid, the sceptical reader might examine pictorial representations derived from a mathematical entity known as the Mandelbrot set[14]. The nonlinear properties of the associated formulae, mapped onto the complex plane, produce an endless and stunning variety of spectacular images (for example see Fig. 6). A literal transcription from the two-dimensional pictorial frame to that defined by frequency and time would prove unsatisfactory. Nevertheless if there is an analogous "fractal" music waiting to be discovered then only digital techniques are likely to find it.

Other more fundamental questions have been touched upon recently by one of France's most experienced and influential computer-composers, Jean-Claude Risset[15]. For example, the access provided by digital systems to what one might call a multidimensional sonic continuum leads to a problem of aural hierarchy: to what extent does musical cognition depend on discrete structures such as scales or rhythms? And once such basic elements of language have been defined, further hierarchical problems arise, as many composers have realised following the musical revolutions that occurred in the early years of this century. Specifically, given completely new possibilities for the synthesis of sound, what are the self-consistent equivalents of the phrase, the counterpoint, the development section or the symphony that sprang naturally from diatonic and chromatic harmony in earlier times?

No doubt the strongest of musical instincts will continue to be required to solve such problems without sacrificing the degree of communication necessary to justify the effort. "It has often been said that music is as much a science as an art;" writes Boulez[16]. "How could these two entities be fired in the same crucible, except by the imagination, that queen of faculties!".

Thanks are due to Pierre Boulez for access to IRCAM, to Andrew Gerzso for introducing me to the workings of *Répons* and for supplying scores and software, to Gerzso and Jonathan Harvey for helpful comments on the article, and to Elaine Gould for her musical calligraphy. □

Philip Campbell is Physical Sciences Editor of Nature.

1. Xenakis, I. *Formalized Music* (Indiana University Press, Bloomington, 1971).
2. Longuet-Higgins, H. C. *Nature* **263**, 646-653 (1976).
3. Erato, STU 71544.
4. Harvey, J. *Computer Music J.* **5**, 22-24 (1981).
5. Dodge, C. & Jerse, T. A. *Computer Music* (Schirmer, New York, 1985).
6. Moorer, J. *Computer Music J.* **1**, 4-37 (1977).
7. De Poli, G. *Computer Music J.* **7**, 8-26 (1983).
8. Mathews, M. V. *The Technology of Computer Music* (MIT Press, Cambridge, Massachusetts 1969).
9. Potacsek, P. & Favreau, E. *Manuel d'Utilisation du Language de Patch* (IRCAM, Paris, 1985).
10. Rodet, X. Potard, Y. & Barriere, J.-B. *Computer Music J.* **8**, 15-31 (1984).
11. Koechlin, O. *et al. IRCAM Rapport Recherche*, 39 (1985).
12. Boulez, P. *Répons* (Universal Edition, 1986).
13. Gerzso, A. *Contemporary Music Rev.* I, 23-34 (1984).
14. Peitgen, H.-O. & Richter, P. H., *The Beauty of Fractals*, (Springer, Berlin, 1986).
15. Risset, J.-C. in *Composers and the Computer* (ed. Roads, C.) (Kaufman, California 1985).
16. Boulez, P. in *Boulez on Music Today*, 143 (Faber and Faber, London, 1971).

Cosmic microwave background anisotropy

Nick Kaiser* & **Joseph Silk†**

* Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK
† Astronomy Department, University of California, Berkeley, California 94720, USA

Current hypotheses for the origin of structure in the Universe lead to predictions of the amplitudes of anisotropies in the cosmic microwave background radiation. The dipole anisotropy is related to density fluctuations on large scales and to other determinations of our motion relative to distant galaxies. Observation and theory are coming tantalizingly close to measuring the elusive anisotropy, or to revealing that our ideas about the origin of galaxies and large-scale structures are in need of substantial revision.

STUDIES of the angular structure of the microwave background radiation play two distinct roles in cosmology. First, there are those anisotropies that are intrinsic to the microwave background. These provide an unsurpassed glimpse of the early Universe, propagating freely to us from an epoch when the Universe was much simpler than it is today. Galaxy clusters had not formed, nor, most probably, had galaxies themselves. Detection of anisotropies in the background radiation would provide a unique measure of the primordial density fluctuations from which the present large-scale structure of the Universe developed. Yet despite continuing and ever-improving measurements, no such anisotropies have been discovered. Here we describe some of the current hypotheses for the origin of structure in the Universe, and the calculations which have been made to predict the amplitude of the microwave anisotropy. A different test of cosmological theories is provided by the dipole anisotropy of the microwave background, which arises from our motion relative to a uniformly expanding 'cosmic frame' defined by the mean state of motion of matter at great distances. We describe this anisotropy as extrinsic, and infer a solar motion which can be used to constrain both the large-scale density fluctuations in which we are embedded and independent determinations of our motion relative to distant galaxies. In the latter part of this review we describe the present status of these studies.

Origin of structure

Large-scale structure originated, according to the primaeval hypothesis, from infinitesimal density fluctuations in the earliest instants of the Big Bang. Inflation[1] has provided a tentative answer to the question of the origin of these fluctuations[2-5]. Quantum fluctuations are amplified to large scales during the exponential expansion phase which characterizes the inflationary epoch. The inflation is driven by vacuum energy during a first-order phase transition which occurs as the energy of matter decays. The universe evolves from the symmetric state of false vacuum, when the fundamental forces are unified, to the asymmetric state of the true vacuum. During the inflationary phase, described by de Sitter space-time, there is no particle horizon. The fluctuations are boosted to a scale-invariant amplitude, up to scales comparable to the radius of the inflated space-time bubble, which greatly exceeds our present horizon, and models can be constructed that yield density fluctuations at horizon crossing of

$$(\delta\rho/\rho)_{\mathrm{H}} \approx 10^{-4} \qquad (1)$$

Values much larger ($(\delta\rho/\rho)_{\mathrm{H}} \approx 1$) would be disastrous for cosmology, producing black holes and large inhomogeneities, and much smaller values would not lead to any galaxy formation. It is remarkable that constraints on the precise value of $(\delta\rho/\rho)$ can be inferred from constraints on temperature fluctuations $\delta T/T$ in the cosmic microwave background. Indeed, all currently acceptable cosmological theories require large-scale structure to have originated from infinitesimal fluctuations at early phases of the Big Bang, and these inevitably leave a signature in the form of temperature fluctuations $\delta T/T$. Whether the amplitude (1) will emerge naturally from a quantum gravity model is not known. Because quantum gravity is poorly understood, most approaches to this problem begin when a classical description first becomes valid, well below the Planck energy of $\sim 10^{19}$ GeV. According to one school of thought, the inflationary potential can be designed to yield the correct $\delta\rho/\rho$, which can then be interpreted as an otherwise unknown particle physics parameter. This process results in a scale-invariant spectrum of gaussian density fluctuations. Another view is that the initial fields that describe the energy content of the Universe are chaotic and highly variable in space and time[6]. Beginning with this initial space-time foam, an anthropic argument is applied to infer that a very special region, with sufficiently regular fields, will inevitably exist and be able to inflate sufficiently to dominate space-time, producing a highly isotropic (and nearly homogeneous) universe. More than one epoch of inflation could have occurred[7], and a wide range of possible primordial fluctuation spectra can arise.

One intriguing consequence is that the primordial fluctuations might be formed exclusively on small scales, far smaller than those of galaxies or galaxy clusters. Galaxy formation can still occur, however, if nonlinear structures develop on stellar mass scales; an amplification process can then produce large-scale structure[8,9]. This happens as follows: massive stars form, explode as supernovae, and sweep up large shells of ambient matter which subsequently fragment on much larger scales, into pieces with the mass of star clusters. If sufficient energy is to be released by dying stars at this stage for the process to repeat, amplifying to larger and larger scales, then formation of structure on galactic mass scales can be achieved. Galactic binding energy, in such schemes, may be said to be of mediaeval, rather than primaeval, origin: it is secondary, having been induced by the explosions altering the distribution of matter.

In the primaeval picture, the excess binding energy of a cluster of galaxies can be traced back to a primordial excess of density in the same co-moving region at the inflationary epoch, whereas in the mediaeval picture this binding energy is imposed at a very late epoch, when the region is well inside the horizon. There is yet another hypothesis, the 'cosmic string' model[10,11], in which the binding energy fluctuations on a given scale are induced by the dynamics of the string network as that scale enters the horizon.

In each of these models there is a link between the very earliest instants of the Big Bang and the large-scale structure of the Universe today. The microwave background fluctuations result from the linear growth regime of the primordial fluctuations,

and thus bypass much of the uncertainty associated with nonlinear evolution. Observations of the microwave background anisotropy provide a unique probe which may help to distinguish between the various models for the origin of galaxies.

Intrinsic anisotropy

Figure 1 shows what one may learn by studying the anisotropy of the cosmic microwave radiation. The microwave photons stream freely to the observer ('here and now') from the surface of last scattering, which in the standard model of the Big Bang occurs in the redshift range $z \approx 1{,}100$–$1{,}500$. The last scattering surface has a finite thickness Δz, which is due to the noninstantaneous recombination of the hydrogen and to the residual ionization level which freezes out. A transmission factor per unit redshift interval, $(d\tau/dz)\exp(-\tau)$, defines the effective profile of the last scattering surface, where $\tau = \int_0^z n_e \sigma_T \, cdt$ is the optical depth to redshift z, σ_T is the Thomson cross-section and n_e is the electron density. This function can be approximated by a gausian which peaks at $z = 1{,}065$, with a gaussian width $\Delta z = 80$ (ref. 12). The corresponding co-moving distance that describes the thickness of the last scattering surface is $\Delta L \approx 7h^{-1}\,\Omega^{-1/2}$ Mpc (where h is the Hubble constant divided by $100\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$).

The finite width of this shell smooths out fluctuations on angular scales $\Delta\theta \leqslant 8\,\Omega^{-1/2}$ arc min. Even larger-scale primordial fluctuations are not necessarily preserved: a source of energetic photons or heat could re-ionize the Universe or increase its ionization over the low post-recombination value ($n_e/n_H \leqslant 10^{-3}$) and thereby shift the last scattering surface towards lower redshift. As the thermal history of the Universe at the relevant epoch ($1{,}000 \geqslant z \geqslant 30$) is uncertain, there is considerable ambiguity in the interpretation of fine-scale anisotropy. This ambiguity is avoided if we search over angular scales exceeding that subtended by the particle horizon at the latest possible epoch of last scattering. No causal process could erase such structure. This angular scale is $\sim(\Omega/z)^{1/2}$ radians, where Ω is the cosmological density parameter. As the minimum redshift at which significant scattering could have occurred ($\tau \geqslant 1$) is ~ 30 if the baryon density parameter $\Omega_b \leqslant 0.1$, we infer that an observational strategy aimed at fluctuations over angular scales of $\geqslant 10°$ would involve the least number of assumptions about the early evolution of the Universe.

The corresponding co-moving linear scale is $\sim 1{,}000$ Mpc. At recombination, a linear scale of 100 Mpc subtends an angle of $\sim \Omega h$ degrees. Observations of the background radiation fluctuations therefore nicely complement studies of galaxy clustering, which provide firm evidence for structure only on scales of up to a few Mpc, in probing the large-scale structure of the Universe. Thus, the interpretation of large-angle anisotropy is independent of the uncertain ionization history, but is crucially dependent on the assumed extrapolation of density fluctuations to large scales by means of equation (1). For small-scale anisotropy the situation is reversed. Regardless of the angular scale considered, the radiation has streamed freely since the Universe was very homogeneous, at which time it was well described by linear fluctuation theory. Thus, in contrast with studies of galaxy formation and clustering[13–16], the complexities of nonlinear evolution are completely bypassed. The detection of anisotropy on any angular scale could provide a unique and direct measure of the primordial fluctuations from which large-scale structure evolved.

Predicting $\Delta T/T$. A variety of calculations predicting $\Delta T/T$ can be found in the recent literature; they differ mainly in the assumed material content of the Universe, the assumed initial fluctuations and the normalization to present-day structure. In order to make sense of the issues here, we first describe one set of calculations, which we refer to somewhat arbitrarily as the standard model, and then describe variations on this theme.

The 'standard' model: cold dark matter with isentropic gaussian fluctuations. The fluctuations in this model have been studied

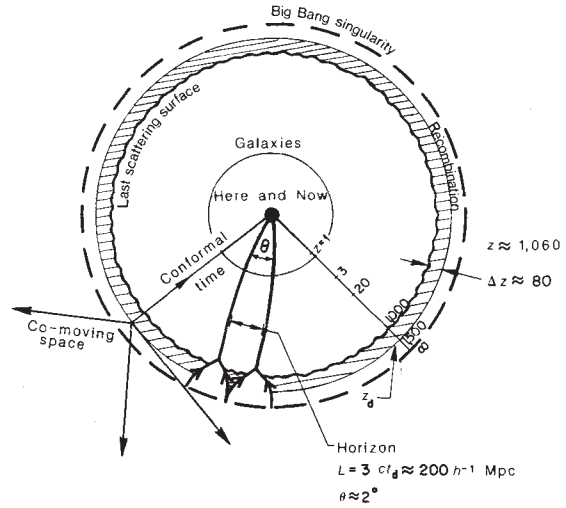


Fig. 1 A co-moving space/conformal time diagram of the Big Bang. The observer ('here and now') is at the centre; the Big Bang singularity is the outermost dashed circle, and the horizon scale is schematically indicated at last scattering.

in detail[17,18]. The Universe is assumed to have three components: a cold, collisionless fluid (this might represent axions, for example); the baryonic gas, initially hot and fully ionized; and the photons and neutrinos, described by a distribution function, with the former being coupled to the gas by electron scattering. The initial state is assumed to be isentropic, which means that, for any two types of particle A and B, the ratio of the number density of As to Bs, n_A/n_B, is a constant independent of position. This terminology arose in the context of a universe containing only radiation and plasma, for which the ratio of the number of photons to the number of baryons n_γ/n_B is a measure of the entropy (per baryon). Such perturbations are sometimes referred to as 'adiabatic' because they can be generated by compressing all of the particles in some region, producing a net density perturbation $\delta\rho(r) \equiv \rho(r) - \rho$. Strictly speaking, one must also specify the initial peculiar velocity field, but this is usually dispensed with by stipulating that the perturbation should contain only growing modes.

A 'gaussian' field is one which can be generated as a linear sum of spatial Fourier components with random phases. Such a field is specified by its power spectrum, which in this case is taken to be the Zel'dovich spectrum, $|\delta_k|^2 = |\delta\rho/\rho|_k^2 = Ak$ (where k is the wavenumber of a Fourier component and A is a constant to be specified by a normalization described below), which results in binding energy fluctuations that are scale-invariant.

With the initial conditions thus specified, the spatial Fourier modes can evolve independently until some time after recombination. This results in final linear fluctuations in the matter and radiation which are now decoupled. The angular dependence of the sky fluctuations is now completely specified. The amplitude, however, remains to be fixed. This 'normalization' is performed by requiring that the density fluctuations be of sufficient amplitude to generate the observed large-scale structure. This is conventionally measured by the integral of the galaxy–galaxy correlation function: $J_3(x) = 3/x^3 \int_0^x \xi_{\mathrm{gal-gal}}(r) r^2\, dr$, where x is a measure of the separation between an arbitrarily selected pair of galaxies. Note that despite the fact that $\xi(r) \gg 1$ on small scales, this integral satisfies linear theory[19]. One evaluates this at a scale where the fluctuations are linear but large enough so that $\xi(r)$ is well known. For example, at $x = 10h^{-1}$ Mpc, the r.m.s. fluctuations are $\sim 50\%$. This normalization assumes that the galaxies are 'fair tracers' of the density field. Although this is a fairly natural assumption, this need not necessarily be the case. If the Universe has closure density ($\Omega = 1$), then it seems that galaxies must be more strongly clustered than mass, at least in the dense clusters where virial analysis is applied to estimate

mass to light ratios. In considering these models, some allowance must be made for this, somewhat reducing the amplitude of $\delta T/T$.

There are two final complications. Uncertainty in the Hubble constant H_0 leads to uncertainty in $\delta T/T$. Specifically, for cold dark matter, a decrease in H_0 leads to an increase in L_{eq} (the radius of the co-moving sphere encompassing the causal horizon at the epoch of equality of matter and radiation densities), and thus to an increase in $\delta T/T$. Similarly, if $\Omega<1$, $\delta T/T$ is increased, approximately in proportion to Ω^{-1}. This is because the Universe becomes curvature-dominated at redshift $\sim 1/\Omega$, and fluctuation growth effectively ceases. One is now in a position where $\delta T/T$ can be predicted. The detailed coupling of matter and radiation must, of course, be followed. This involves solving the perturbed Boltzmann equation for the photon intensity today[20,21]. Rather than give any details here, we will simply indicate in a qualitative manner the magnitude of the various terms that contribute to $\delta T/T$.

On small scales, corresponding to a co-moving scale that subtends the thickness of the last scattering surface at ~ 10 Mpc, there are adiabatic temperature fluctuations. These were first predicted in 1967[22,23], and would be of order $(1/3)\delta\rho/\rho$ were recombination instantaneous and complete. Their angular scale is typically 5–10 arc min, and fluctuations of smaller angular scale do not survive recombination. The amplitude of these small-scale fluctuations is reduced, because of the finite thickness of the last scattering (LS) surface, to $\delta T/T \approx (\delta\rho/\rho)_{LS}/30$. On larger scales, the random motions of fluctuations induce Doppler shifts. These dominate $\delta T/T$ at co-moving scales $d \approx L_{rec}$ (the radius of the co-moving sphere encompassing the causal horizon at recombination) at $\sim$50–100 Mpc, corresponding to angular scales of 1–2°, and are of order $\delta T/T \approx (\delta\rho/\rho)_{LS}(d/ct)_{LS}$, where c is the speed of light and t is the time elapsed since the Big Bang. On the largest scales, the gravitational potential difference between last scattering and now generates gravitational redshifts, of order $\delta T/T = (\frac{1}{3})(\delta\rho/\rho)_0(d/ct)_0^2$. These fluctuations, also first predicted in 1967[24], are contributed by co-moving scales d of 10^3–10^4 Mpc, and dominate over angular scales greater than a few degrees[18].

The result of all this is a gaussian sky pattern (the gaussian nature of the fluctuations being preserved in linear theory), with roughly equal power in each logarithmic interval of angular wavenumber. The r.m.s. amplitude is a function of global cosmological parameters such as Ω, h and Ω_b, and constraints can be placed on these by comparison with the stringent small-scale anisotropy limits. High-density models are generally consistent with these limits, but low-density models can be excluded. There are two reasons for this: first, there is reduced growth of $\Delta\rho/\rho$ at late times if $\Omega \ll 1$, and second, the fluctuations appear at a more favourable angular scale. The results we have described were obtained numerically. The anisotropy on large scales, much greater than the horizon size at decoupling, for these isentropic fluctuations can be calculated analytically[24] (see ref. 25 for an application to cold dark matter). The effect is analogous to the usual gravitational redshift effect, with photons being redshifted (blueshifted) if they were last scattered in overdense (underdense) regions. The amplitude of $\Delta T/T$ is one third of the newtonian peculiar potential ϕ evaluated at the point of emission.

In these calculations it is assumed that the Universe is not re-ionized after recombination. There is some justification for this in that, for most reasonable choices of parameters, the spectrum of density fluctuations is such that very little matter forms nonlinear structure at a sufficiently early epoch to plausibly re-ionize the Universe[17,18]. We now describe some variations on this 'standard' model.

Hot dark matter. An interesting alternative is to replace the cold dark matter (axions, photinos, or the like) with hot dark matter such as massive neutrinos. The main effect of this is to erase post-recombination density fluctuations on scales smaller than that of superclusters (the angular dependence of $\Delta T/T$ being largely unaffected because the $\Delta T/T$ fluctuations on small scales are in any case smeared out)[17,26]. The normalization to J_3 (for galaxies) is now questionable, because galaxy formation is a secondary phenomenon in a neutrino-dominated universe, arising from pancake fragmentation[27]. An alternative normalization is to require that 'pancake' collapse should occur sufficiently early to form quasars and the most distant galaxies. For a neutrino mass of $\sim$30 eV, the predicted anisotropy is compatible with the current upper limits.

Isocurvature fluctuations. An alternative to the isentropic fluctuations discussed so far is the 'isocurvature' mode. These fluctuations are, in a sense, orthogonal to the isentropic modes because, while the latter have fluctuations in the total energy (and therefore in space curvature) but no variation in the relative abundances of the various particles, fluctuations can be realized by spatially varying the equation of state on some initial $t=$ constant hypersurface in a universe which was hitherto absolutely homogeneous. This process conserves energy locally, so any excess in one species is balanced by a deficit in the others. It is the constancy of the spatial curvature on this initial hypersurface which gives rise to the terminology employed here. These fluctuations are not, strictly speaking, isothermal, but are often referred to as 'entropy' or 'isothermal' fluctuations.

Isocurvature fluctuations in a universe dominated by cold dark matter have recently been investigated[28,29]. The small-scale anisotropy is very similar to that arising from isentropic fluctuations. The large-scale anisotropy, however, is about an order of magnitude larger. This increase is a consequence of two factors: the first is that the spectrum of density fluctuations is flatter than in the isentropic case, because no sub-horizon growth occurs in the radiation era. Thus, with the same normalization, the value of $\Delta\rho/\rho$ at large scales is about twice as large.

The origin of the other factor can be understood as follows. Imagine that, at some early time $z \gg z_{eq}$, we impose a positive perturbation to the axion density $\Delta\rho_A/\rho_A = \Delta n_A/n_A = \varepsilon$ in some large spherical region destined to enter the horizon at some time after z_{EQ}. We must also impose a smaller compensating negative perturbation in the radiation density, $\Delta\rho_\gamma/\rho_\gamma = -(\rho_A/\rho_\gamma)\varepsilon$, so that ρ_{TOTAL} is unperturbed. When this region enters the horizon, the density of the radiation will be negligible, having been redshifted, so we will have a positive perturbation to the proper mass enclosed: $(\Delta M/M)_{proper} = \Delta n_A/n_A = \varepsilon$. However, from considerations of causality, we know that the gravitational mass as registered by an external observer is unperturbed. There is no paradox here; the perturbed region must now be sitting in a potential well of depth $\phi = \varepsilon$ so that $M_{grav} = M_{proper}(1-\phi)$ is unchanged. If the perturbation were isentropic then we would see the Sachs–Wolfe term, $(\Delta T/T)_{SW} = -\phi/3 = -\varepsilon/3$. However, this perturbation is not isentropic: we initially increased the number of axions but the number of photons was almost unaltered. Thus, relative to an isentropic state, we have a negative entropy perturbation, $\Delta n_\gamma/n_\gamma = -\varepsilon$, which gives rise to a temperature fluctuation $(\Delta T/T)_{entropy} = 4(\Delta n_\gamma/n_\gamma)/3 = 4\varepsilon/3$. We see then that the usual Sachs–Wolfe term is augmented by an entropy perturbation of the same sign but four times larger, and the net result is to increase $(\Delta T/T)_{TOTAL}$ by a factor of $\sim$5. With the conventional normalization, the predicted anisotropy is incompatible with the limits of Fixen *et al.*[30].

The re-ionized Universe. If re-heating were to occur, the Universe would become transparent over approximately one expansion time at the mean redshift of last scattering. The photons we see would then arrive from a broad range of redshifts, $\Delta z \approx z_{ls}$. The temperature anistropy due to the large-scale structure in such a re-ionized universe is smaller than that predicted to arise in a un-ionized universe. This is because, along any line of sight through the cosmic photosphere, we will see fluctuations due to many regions adding incoherently and this results in statistical cancellation.

The dominant source of anisotropy in a re-ionized universe

is the peculiar motion of the last scatterers. The calculation of the anisotropy is greatly simplified by two approximations which are expected to hold to high accuracy. First, at the epoch of last scattering, the density fluctuations that give rise to the currently observed large-scale structure were still in the linear regime, and second, the dynamical effect of the radiation on the matter was almost certainly negligible. Hence we need only calculate the effect of the matter on the radiation.

The statistical cancellation due to the projection of density fluctuations along the line of sight depends on the statistical nature of the density perturbation field $\delta(k)$. If this density field was gaussian (that is, if the Fourier components $\delta(k)$ had random phases), then the temperature fluctuations on the sky, $\Delta(\theta)$, will also be a gaussian field. The power spectrum of this field, Δ^2, is related to the power spectum of density fluctuations by $\Delta^2 \propto k^{-4}\delta^2(k)$, where the constant of proportionality depends on the ionization history of the Universe[31]. A consequence of the fact that the scale of the fluctuations is much smaller than the thickness of the photosphere is that the fluctuations $\Delta(k)$ at some angular wavenumber k are determined solely by the value of δ^2 at the corresponding spatial wavenumber.

A useful measure of the amplitude of the temperature fluctuations at angle θ is $(\delta T/T)_\theta \approx \sqrt{k^2\Delta^2(k)}$, where $k \approx 1/\theta$. Similarly, the amplitude of the density fluctuation on scale λ is $(\Delta\rho/\rho)_\lambda \approx [k^3\delta^2(k)]^{1/2}$. Thus, to order of magnitude, we have $(\Delta T/T) \approx (H\lambda)^{5/2}(\Delta\rho/\rho)$ (where H, λ and $\delta\rho/\rho$ are evaluated at z_{ls}). Equivalently, we can write $(\Delta T/T) \approx (1/\sqrt{N})\phi_\lambda$, where ϕ_λ is the gravitational potential fluctuation on scale λ and N is the number of regions of size $\sim\lambda$ seen along a line of sight. One can infer that the anisotropy due to fluctuations of a given scale λ is smaller for smaller z_{ls} (as N is then increased), and ϕ_λ is approximately equal to the square of the velocity dispersion of the final nonlinear structures produced. As this velocity dispersion appears to be a non-decreasing function of mass (and therefore of λ), the result tells us that the greatest anisotropy is generated by the largest structures we see today. If galaxies fairly trace the mass distribution, then the anticorrelation seen in the redshift survey suggests that the power spectrum falls rapidly for $k \leq 2\pi/60h^{-1}$ Mpc. These fluctuations would generate $\Delta T/T \approx (3\text{-}6)\times 10^{-6}$ on an angular scale of $\sim 1°$. (We have here assumed $\Omega \approx 0.2$ in accord with the assumption that galaxies trace the mass). In a high-density universe ($\Omega \approx 1$), the amplitude would be smaller by a factor of ~ 3 and the anisotropy would appear at an angular scale of ~ 10 arc min.

On larger scales, the amplitude of $\Delta T/T$ depends on the extrapolation of the power spectrum. If this is assumed to be a power law with $\delta_k^2 \propto k^n$, then $\Delta T/T \propto \theta^{(-n+2)/2}$ and would increase with angular scale if $n \leq 2$. The estimate we have given above is the minimal anisotropy implied by the structure observed today, and it does not require any extrapolation of the power spectrum to large scales. The predicted anisotropy is well below the current upper limits on the relevant scales.

So far we have not specified the source of energy which re-ionizes the Universe. An interesting possibility is that these pre-galactic sources of radiation may, if they are sufficiently inhomogeneous, generate the large-scale structure[32]. A constraint on such theories is that, in addition to the 'Doppler scattering' fluctuations we have discussed above, the radiation released would probably have been degraded to microwave frequencies and would appear anisotropic on a scale of a few degrees, this being the angular size subtended by the width of the photosphere.

This effect was investigated in ref. 33. The assumptions are that a population of sources turn on and burn at redshift z_*, that these sources are inhomogeneous on some scale λ_* and are uncorrelated on larger scales (note that there is no assumed initial density inhomogeneity), and that these sources produce a substantial fraction of the microwave background. The diffusion or streaming of this radiation away from the sources then generates density fluctuations by the pressure of the radiation of the gas. This process can transfer momentum, and generate density fluctuations, on scales up to the radiation Jeans length, which is comparable to the scale of the largest structures we see today. The redshift of burning was taken[33] to be $z_* \approx$ 200–500; if z_* is lower than this, then radiation drag becomes inefficient, and for higher z the luminosity of the sources must exceed the Eddington luminosity. This process could generate the observed large-scale structure if the initial scale of radiation inhomogeneity is of the order of a galactic mass scale. The transfer of radiation smooths out the initially inhomogeneous radiation on scales up to the horizon scale at last scattering. The final microwave background anistropy predicted (assuming thermalization of the radiation by dust) appears at an angular scale of a few degrees and is comparable to the present upper limits on this scale.

On larger scales the temperature anisotropy decreases as $\Delta T/T \propto \phi^{-1}$, and so any quadrupole anisotropy would be very small. This $\Delta T/T \propto \phi^{-1}$ behaviour at large angles is the hallmark of spontaneous theories of formation of structure. Most 'primordial' theories of large-scale structure invoke a Zel'dovich spectrum which would generate large-angle $\Delta T/T$ fluctuations which are scale-independent. Thus observation of large and comparable quadrupole, octopole and higher-order moments would rule out spontaneous generation theories. Unfortunately, our Galaxy gives rise to anisotropy, with significant low-order multipoles. It may then be that one would be able to detect the Zel'dovich anistropy only on intermediate angular scales, where one could test for a dependence of $\Delta T/T$ on galactic latitude.

Anisotropy from cosmic strings. All of the mechanisms we have discussed so far give rise to gaussian fluctuations, either by virtue of the assumed initial conditions or, as in the re-ionized universe, where we have many fluctuation regions adding incoherently, so that the central limit theorem should apply. An interesting alternative is supplied by the 'cosmic string' hypothesis[10,11] (for a review of this subject see ref. 34). The perturbations provided by cosmic strings are inherently non-gaussian and result in step-like discontinuities snaking across the microwave sky[35-37]. The pattern is like that arising from a Zel'dovich spectrum, scale-invariant on large scales. In this picture, one would naturally expect the Universe to be re-ionized, so the theory is relatively immune to the constraints from small-scale anisotropy. In order to detect the distinctive features predicted in this model, one would need to scan a region of sky at least $\sim 10°$ square with reasonably good resolution.

ΔT/T in the explosive model. Explosive amplification[8,9] begins with rare seeds, either massive objects ($\sim 10^6$ solar masses) or possibly galaxies themselves, and results in a foam-like distribution of galaxies: swept-out voids separate thin shells, sheets and ridges where shells intersect, and there most of the galaxies form. The typical void size can be up to ~ 5 Mpc in diameter if sufficient explosive energy is available. Although this falls short of the observed voids in the galaxy distribution, which are ~ 25 Mpc across[38], the qualitative resemblance of the bubble-like structure of the galaxy distribution to that predicted in explosive amplification is tantalizing, and this model is a serious contender for explaining galaxy formation.

In fact, the voids are not completely empty, but contain hot shocked gas at a temperature of $\geq 10^8$ K. In the pancake fragmentation model, the physics of galaxy formation is very similar to that in the Ostriker–Cowie–Ikeuchi model[8,9]. Pancakes collapse to form thin, dense sheets of gas which fragment into galaxies, leaving behind large regions of tenuous hot gas. This hot gas can result in observable fluctuations in the cosmic microwave background, by means of the Sunyaev–Zel'dovich effect[39]. Microwave photons Compton-scatter against the hot electrons and acquire a perturbation in temperature, in the Rayleigh–Jeans part of the spectrum, of $\Delta T/T \approx -(2kT/m_e c^2)\tau_{es}$, where k is the Boltzmann constant, m_e is the electron mass and τ_{es} is the scattering optical depth. $\Delta T/T$ for

an individual pancake is proportional to the pressure, which is constant in the shocked region and depends only on the co-moving wavelength of the pancake and the epoch of pancake collapse. Although the distortion from an individual pancake at $z \approx 1$ is small ($\Delta T/T \approx 10^{-7}$), the contribution from the many ($N > 100$) pancakes along a typical line of sight increases the variance of the temperature fluctuations by a factor of $N^{1/2}$ between different lines of sight. Hogan[40] performed a similar calculation for the explosive galaxy formation model. He found that if the observed large-scale structure as measured by the galaxy autocorrelation function is generated explosively, small-scale anisotropy comparable to the Uson-Wilkinson limit[41] is produced. Vishniac and Ostriker[42] generalized this argument further, by noting that if the binding energy of the gaseous components of newly bound structures is thermalized, regardless of its origin (whether explosive or gravitational), then temperature fluctuations of amplitude $\sim 10^{-5}$ result on sub-arc-minute scales. Primordial fluctuations are greatly diminished on these scales due to the smearing associated with the finite thickness of the last scattering surface, and such secondary fluctuations are expected to dominate.

Searches for intrinsic anisotropy. Since the discovery of the microwave background radiation[43], many attempts have been made to search for angular anisotropy. These searches have used a variety of techniques to probe a range of angular scales spanning more than three orders of magnitude. It is often convenient, in the discussion of temperature anisotropies, to decompose the sky temperature pattern into spherical harmonics; these provide a complete and orthogonal set of modes. As we have mentioned, the $l = 1$ mode is probably dominated by our locally generated peculiar velocity, which may also generate a small $l = 2$ component. Any other anisotropies must be intrinsic. Observations at different angular scales suffer from different problems and require different observing strategies; it is useful to distinguish between 'large' angular scales for which $l \leqslant 10$, 'intermediate' angular scales with $l \approx 10$–100, and 'small' angular scales with $l \geqslant 100$.

For large and intermediate angular scales it is necessary to raise the observing apparatus above the bulk of the atmosphere. Anisotropometers have been deployed from balloons[30,44-46], from U-2 aircraft[47] and from a satellite[48]. These observations have typically employed beam-widths of a few to ten degrees and have obtained coverage of a significant fraction of the whole sky. Initially there was much emphasis on the possibility of a quadrupole anisotropy, and observers obtained estimates of the quadrupole tensor components from their data. These observations also contain useful information about higher-angular-frequency fluctuations on scales down to the beam-width. Although there have been claims of positive detections of quadrupole anisotropy at levels significantly above that of the noise, none of these have withstood the test of time, and the current upper limit from the Soviet RELIC experiment on any $l = 2$ mode is $\Delta T/T \leqslant 3 \times 10^{-5}$ with 95% confidence. The limit on the $l = 4$ mode is 5×10^{-5} (95% confidence). These limits are much smaller than the dipole amplitude, and this fact gives some support to the belief that most, if not all, of the dipole anisotropy is extrinsic or local, since one would expect any intrinsic dipole anisotropy to be of similar amplitude to the other low-order moments.

On intermediate angular scales there has also been one claimed positive detection of anisotropy[44], although few details of the method of analysis were given and no confidence level was quoted. Other workers have quoted upper limits of $\Delta T/T \leqslant 3 \times 10^{-5}$. More recently, an upper limit[49] of $\Delta T/T \leqslant 5 \times 10^{-4}$ (95% confidence) has been set at a beam-width of 2°. An upper limit of 5×10^{-5} (95% confidence) has been set over ~6–12° in a recent dedicated experiment at 3 cm wavelength (R. D. Davies *et al.*, in preparation). At small angular scales the emission from the atmosphere is less of a problem, and searches for anisotropy on scales of a few arc min have been made using conventional ground-based radio telescopes. The comparison of the results from these various observations, and their relation to quantities which can easily be predicted from theoretical models, are hampered by the variety of observing strategies employed. In order to achieve the high sensitivity required, it is necessary to perform a differential measurement; many different choices of beam geometry have been used. One possibility would be to allow the beam to drift across the sky and measure the r.m.s. fluctuations in the receiver output. This would then tell us the r.m.s. fluctuation in the sky temperature after this has been convolved with the beam pattern. In practice, however, a more common strategy is to repeatedly switch between two or more patches of sky and subtract the signal. If only two patches are observed, the output is simply the autocorrelation function of the (convolved) sky brightness. If, as is usual, the beam throw is comparable to the beam-width, then the output is a useful measure of the fluctuations on this angular scale. Such an apparatus would also have some sensitivity to fluctuations on much larger angular scales, because the apparatus measures the gradient of long-wavelength fluctuations. The most stringent current upper limits to small-scale anisotropy are those obtained by Uson and Wilkinson[41], using a double beam subtraction technique. They quote an upper limit $\Delta T/T \leqslant 2 \times 10^{-5}$ at $\theta \approx 4.5$ arc min, again with 95% confidence. This type of measurement is even less sensitive to large angular fluctuations because the beam geometry measures the second derivative of the sky pattern at large angles. Finally, we mention some searches for anisotropy on 'very small' angular scales (a few arc s), using the Very Large Array[50]. The sensitivity is somewhat poorer than that obtained on larger angular scales, with upper limits of 10^{-3}, 8×10^{-4} and 5×10^{-4} (95% confidence) at angular scales of 18, 30 and 60 arc s, respectively. These results have now been superseded by new, deeper survey at 6 cm, which improves on these limits by factors of 4 and 8 at 18 and 60 arc s, respectively[51]. Similar results have been obtained independently by H. M. Martin and R. B. Partridge (in preparation).

We also expect the microwave sky to display some linear polarization. To date, polarimetry searches have been carried out only at large angular scales[52,53], these searches having been stimulated by the suggestion[54] of a quadrupole polarization pattern in an anisotropic universe. Re-ionization can enhance this effect[55]. Polarization with a much smaller coherence angle will also arise due to inhomogeneity. The amplitude of the polarization pattern is of the order of the temperature anisotropy on a scale corresponding to the width of the last scattering shell[56]. If temperature anisotropy is ever detected, then polarization studies will provide an important test of the interpretation of such a result. It may be that temperature anisotropy searches will eventually be limited by confusion from faint sources. If this is the case, then polarimetry may provide the best way to view the primordial fluctuations.

Confidence intervals

The figures quoted above for upper limits on sky temperature fluctuations are a highly reduced form of the initial data, which typically consist of a set of N temperature difference estimates D_i together with estimates σ_i of the experimental uncertainty associated with each D_i. Before we consider the implications of the quoted upper limits it is important to understand the limitations of these statistics. To this end we will briefly discuss the method by which these confidence limits are arrived at and highlight some of the associated problems.

We will consider, for illustrative purposes, the simplified, though not entirely unrealistic example in which the D_i are measured in widely separated fields, and so can be considered to be statistically independent, and in which the instrumental variances are all equal: $\sigma_i^2 = \sigma_0^2$. In this case, the commonly adopted procedure[57] is to calculate the statistic $\chi^2 \equiv \Sigma D_i^2/\sigma_o^2$, and to ask, first, if there is firm evidence for fluctuations in excess of that produced by the receiver. Assuming that there is

not, we then ask what is the value of σ^2_{ul} such that, if the true sky r.m.s. temperature difference is $\sigma_{sky}=\sigma_{ul}$, a value of χ^2 as low as that observed will occur only a fraction $1-c$ of the time. The number of σ^2_{ul} is our upper limit on the sky variance at confidence level c.

One problem with this method is that it requires a rather precise estimate of the instrumental variance σ_0 (ref. 58). For reasonably large N, the central limit theorem tells us that χ^2 will have a gaussian distribution with mean $N(\sigma_0^2+\sigma^2_{sky})/\sigma_0^2$ and variance $2N(\sigma_0^2+\sigma^2_{sky})^2/\sigma_0^2$. The upper limit is then $\sigma^2_{ul}=\sigma_0^2\{(\chi_0^2/N)/(1-\alpha\sqrt{2}/N)-1\}$, where χ_0^2 is the value actually observed and α is the number of standard deviations associated with the confidence level c (for example, $\alpha=1.65$ for $c=95\%$). χ^2/N will be close to unity, $\chi^2/N=(1+O(N^{-1/2}))$, and so $\sigma^2_{ul}\approx\sigma_0^2N^{-1/2}$. But the statistic χ^2 is inversely proportional to the estimate of σ_0^2, so σ^2_{ul} to be reliable, any systematic functional errors in σ_0^2 must be much smaller than $N^{-1/2}$. If we 'generously' overestimate σ_0^2, we will obtain much too low a value for σ^2_{ul} and we will be in danger of excluding the true hypothesis.

Another problem is that, by our definition of σ^2_{ul}, we will reject the true hypothesis a fraction $1-c$ of the time. Thus, if 20 upper limits at 95% confidence are found in the literature than the 'best' upper limit is likely to lie below the true value. The actual situation will be worse than this if there is any selection pressure to favour the appearance of 'good' results. This might occur if there is any tendency, either of a voluntary or involuntary nature, not to publish results which are deemed to be uninteresting compared to other published results. Such selection effects cause the number of independent samples of event space to exceed the number of published results. The rate of sampling of event space can also be boosted by combining sets of data from different runs, or if the decision to extend or curtail the experiment is made after any data have been collected.

If for these or other reasons, the value of σ^2_{ul} obtained has any influence on the publication of results, then the value of these constraints for rational decision making is debased. Ultimately, the only practical way to ensure that results are free of such bias is for the decision to publish or not to be made before the experiment is performed (with a blank space in the manuscript for σ^2_{ul} to be inserted, of course). It is after all only rational that the criterion for acceptance of such a paper should be the sensitivity of the apparatus, rather than the random number σ^2_{ul} which results at the end of the day.

Unreasonably restrictive upper limits may then occur by chance, or they may even be selected for. However, one should not simply accept this as a painful fact of life. Our procedure is guaranteed to reject the true hypothesis 5% of the time (for 95% confidence limits), but in any particular instance the data may give some indication as to whether this is such a case. For example, if $\sigma^2_{sky}=0$, then negative upper limits will occur. In such a case, we can be certain that we have been 'lucky'. There are other cases in which a positive limit results on the variance but in which one would be reluctant to reject the hypothesis that $\sigma^2_{sky}\geqslant\sigma^2_{ul}$. These are those cases in which, while our statistic indicates that the result is unlikely under the hypothesis $\sigma^2_{sky}>\sigma^2_{ul}$, the result is not significantly more likely under any alternative hypothesis. The problems associated with various statistical procedures for treating null results will be discussed elsewhere (A. N. Lazenby and N.K. in preparation), but for the present we note that the hallmark of such a spurious result is an unusually low value of χ^2. Visual inspection of the data presented in ref. 41 suggests this may have been the case in this instance. This is unfortunate because, taken at face value, this result rules out some interesting hypotheses.

Implications

Intermediate and large scales. The most immediate implication of these negative results is that we inhabit a universe that is currently uniform to a high degree of precision, at least on scales greater than the horizon size at last scattering. This inference is particularly certain because considerations of causality assure us that there is no mechanism which can render the Universe isotropic on scales exceeding the horizon scale at last scattering. Although we can observe only that region of space which is inside our present horizon, we can invoke the copernican principle to exclude such possibilities as that we inhabit an inhomogeneous universe which just happens to be isotropic about the Earth. We can also infer that the Universe is very close to homogeneous on scales much larger than the present horizon size[59]. If our universe is nearly homogeneous today, then it must have been prepared in a state with an extremely small amount of 'growing-mode' inhomogeneity present at early times; this preparation may have involved an inflationary phase[1]. The process which prepared the Universe may also have imprinted small departures from perfect homogeneity which could later grow to produce galaxies and other structure in the Universe. If this is the case, then it is most natural that the fluctuations should have curvature (specific binding energy) fluctuations which are independent of scale. Were this not the case, then it would be necessary to impose a special scale (that of curvature nonlinearity) in the initial conditions and this scale would not be many orders of magnitude different from the scale of galaxies. It is hard to see how the microphysics of the early Universe could produce such a scale.

On large scales the spectrum maintains its initial form. We can estimate the level of anisotropy which should be observed on intermediate and large angular scales, because the temperature fluctuations should be at least as large, to order of magnitude, as the specific binding energies of the largest structures we see today. Clusters have internal velocity dispersions of $v\approx1000\ \mathrm{km\ s^{-1}}$ and so have specific binding energy $\phi\approx(v/c)^2\approx10^{-5}$. The binding energies of superclusters are thought to have similar magnitude and so we predict $\Delta T/T\approx10^{-5}$. This is not much smaller than the current upper limits on these angular scales and so it seems that, with a modest improvement in the observational situation, we have the hope of either ruling out this picture or confirming the presence of these scale-free temperature fluctuations.

To obtain a quantitative constraint on $\Delta\rho/\rho$ on these scales, one can use the relation of Sachs and Wolfe[24] for the temperature fluctuation induced by a plane-wave density ripple. A convenient parameter to specify the amplitude of the perturbation is ϕ, the newtonian specific binding energy. The result of Sachs and Wolfe is simply that $\Delta T/T=\phi/3$. We have plotted the constraint on $(\Delta\rho/\rho)_\lambda$ using this relation, and the observed upper limits are shown in Figure 2. This constraint assumed that the initial fluctuation spectrum was initially isentropic.

Small scale. Observations on intermediate and large angular scales provide the best constraint on the amplitude of very-long-wavelength density fluctuations. On these scales, however, there is very little direct evidence for clustering of matter. On scales of $\leqslant10h^{-1}$ Mpc we see strong clustering, at least of luminous material. The importance of small scale anisotropy studies is that they probe those length scales on which we actually see clustering, and so they provide a more direct test of theories of galaxy formation. Unfortunately, however, the interpretation of smallscale anisotropy is model-dependent: in order to apply this test one must first choose the nature of the initial fluctuations (for example, 'isothermal' or 'adiabatic'), assume some spectrum (such as a power law) and then propagate these fluctuations from some sufficiently early time until after the epoch at which the matter and radiation decouple. This linear evolution depends on the properties of the major constituents of the Universe at the time, so the abundances and masses of these particles must be specified. Up to this point the amplitude of the initial fluctuations is arbitrary. The next, and most uncertain, step is to normalize the calculation by requiring that the density fluctuations should have grown to the presently observed amplitude by today. One can then compare the predicted anisotropy with observations and thereby constrain the initial parameter space.

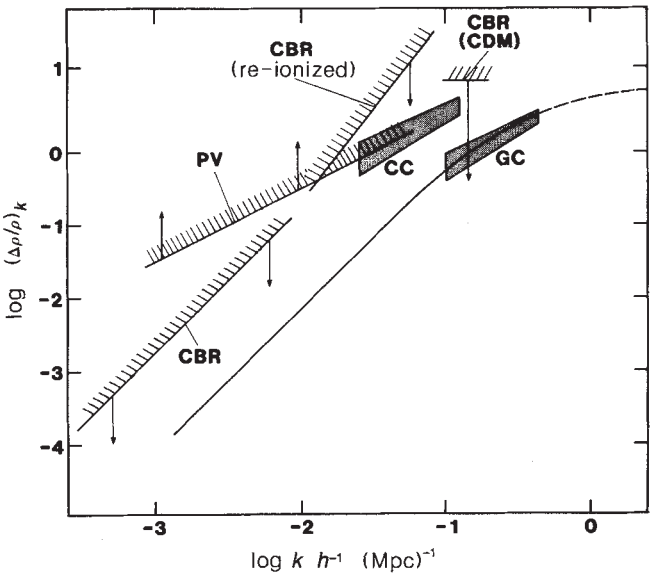


Fig. 2 Constraints on the power spectrum of density fluctuations in an Einstein–de Sitter ($\Omega = 1$) universe. $(\Delta\rho/\rho)_k \equiv k^{3/2}\delta_k/\sqrt{2}\pi$. The contribution to the r.m.s. density fluctuations per logarithmic interval of k, is plotted versus k, where $k(\mathrm{Mpc}^{-1})$ is comoving wave number, $\delta_k = \int d^3k\, e^{ik.x}\, \Delta\rho(x)/\rho$, and x labels the spatial coordinates of an arbitrary point. Specifically, the mass fluctuation within a sphere of radius x is $\delta M/M = \int k^2\, dk |\delta_k|^2 W(kx)/\sqrt{2}\pi$, where the window function $W(kx) \approx 1$ for $kx \leqslant 1$ and $W(kx) \approx (kx)^{-4}$ for $kx \gg 1$. CBR, large- and intermediate-scale microwave anisotropy; we have taken $\Delta T/T \leqslant 3\times10^{-5}$ for $\theta \leqslant 10$ arc min and $\Delta T/T \leqslant 10^{-4}$ for $\theta \approx$ a few degrees. CBR(CDM), small-scale anisotropy limits in a universe dominated by cold dark matter. CBR(re-ionized), anisotropy limits if the intergalactic medium is re-ionized at large redshift and secondary fluctuations are generated as discussed in the text (see discussion of 'the re-ionized universe'). PV, constraint from peculiar velocity; power spectrum of galaxy clustering; CC, power spectrum of cluster clustering; CDM, theoretical preduction for density fluctuations in a cold dark matter model with $h = 0.5$. In a low-density universe ($\Omega = 0.2$), large- and intermediate-scale anisotropy limits are ~3 times lower; peculiar velocity limits are ~3 times higher; small-scale anisotropy limits are ~10 times lower.

In recent years, there have been interesting theoretical and observational developments concerning the small-scale anisotropy. On the theoretical side there has been a rise in popularity of models in which the bulk of the gravitating matter is in a form which is otherwise very weakly coupled to other forms of matter. These models generate a much smaller level of anisotropy than models in which the density is dominated by baryons. Meanwhile, the observational limits have improved, and now not only rule out baryon-dominated models (or at least those with 'adiabatic' fluctuations) but also place useful constraints on models involving weakly interacting 'dark' matter.

The normalization that has been used in these calculations is that the density contrast in spheres of radius $8h^{-1}$ Mpc should have unit variance today. This normalization is chosen because the fluctuations in galaxy counts in spheres of this size have unit variance and because it is thought that fluctuations on these scales should be described fairly accurately by linear theory, and so should be accessible to theoretical calculation. Ideally, it would be desirable to normalize on even larger scales where linear theory would be even more accurate, but unfortunately the estimates of fluctuations in galaxy counts on larger scales become very uncertain.

With this choice of normalization, the small-scale anisotropy limit allows one to exclude 'cold dark matter' models with $\Omega = 0.2$ (see the constraint labelled CBR (CDM) in Fig. 2). This conclusion is particularly interesting because this is the value for Ω implied if galaxies fairly trace the mass distribution. If the value of Ω is significantly greater than 0.2, then galaxies are, for some reason, more strongly clustered than the matter. If this is the case, then it is inconsistent to use the conventional normalization. Inclusion of this effect would lower the predicted $\Delta T/T$ still further in addition to the approximate Ω^{-1} scaling (by a factor of ~2)[60–62]; to detect such tiny fluctuations would require a great improvement over the current upper limits.

The possibility of ruling out a high-density Universe dominated by cold dark matter from small-angle anisotropy is therefore fairly remote. Large-angular-scale searches present a more hopeful prospect, although in this case the interpretation depends crucially on the assumed slope of the primordial spectrum of fluctuations. As we have mentioned, there are reasons to prefer the Zel'dovich, scale-free curvature spectrum. To rule out, or to confirm at the several-sigma level, the existence of the consequent scale-free temperature fluctuations would be a significant achievement.

Extrinsic anisotropy

The microwave sky has now been almost completely mapped at an angular resolution of a few degrees. The only convincing feature which has been observed to date is the 'dipole' anisotropy. Averaging together the two most precise of the recent measurements yields a dipole anisotropy with amplitude $\delta T/T = 1.2\times10^{-3}$. The usual interpretation of this anisotropy is that the Sun is moving at a velocity $v_{\mathrm{solar}} \approx 380\ \mathrm{km\ s^{-1}}$ towards an apex ~45° away from the Virgo cluster. The reason for this interpretation, rather than as a combination of extrinsic and intrinsic anisotropy, is that this dipole is at best an order of magnitude larger than the quadrupole or other low-order spherical harmonics on the microwave sky, and such an intrinsic pattern seems highly implausible. It may be possible to test this interpretation if the intrinsic quadrupole is really very small because, in addition to the dipole of amplitude v/c, our motion will also result[63] in an aligned quadrupole of amplitude v^2/c^2.

This measurement of the solar motion provides us with a constraint on large-scale density fluctuations because, if these induce an r.m.s. amplitude of $\gg 300\ \mathrm{km\ s^{-1}}$, we must live in a very special position[64,65]. Unless there is some anthropic argument to favour the existence of observers in regions which seem to be special only in so far as they have an unusually low dipole anisotropy, we can exclude such hypotheses as unlikely. These hypothetical large-scale density fluctuations should be well described by linear theory, and we then have $v_{\mathrm{rms}} \approx HL(\Delta\rho/\rho)_{\mathrm{rms}}$, for fluctuations on length scale L. If we assume for the moment that the density fluctuations are a gaussian process, then the large-scale velocities will also be gaussian-distributed. We then have an upper limit at 95% confidence of $v_{\mathrm{rms}} \leqslant 3v_{\mathrm{solar}}$ as only ~5% of all observers have a velocity less than one third of the r.m.s. value. We have assumed a gaussian distribution, but this result is not very sensitive to the form of the tails of the actual distribution because we are calculating the probability of very small values of v/v_{rms}. The result would be misleading, however, if the probability $P(v) \not\approx v_{\mathrm{rms}}^{-3}$ for $v \ll v_{\mathrm{rms}}$, as would be the case if a small fraction of observers have $v \gg v_{\mathrm{rms}}$ while most are nearly at rest. Note that we have ignored contributions to v_{solar} from small scales which are nonlinear ($\Delta\rho/\rho \geqslant 1$). This is legitimate if we have no detailed knowledge of the form of these velocities, as the inclusion of small-scale $\Delta\rho/\rho$ will only broaden the distribution $P(v)$. But as we do have some knowledge of the Sun's motion in the Galaxy and the Galaxy's motion in the Local Group, we should incorporate this information. In doing so, we will be able to 'filter out' the nonlinear velocities and concentrate on those linear velocities of interest.

The first step in this process is to allow for the rotation of the Galaxy. This results in a larger velocity, $v_{\mathrm{galaxy}} \approx 550\ \mathrm{km\ s^{-1}}$, and therefore a weaker constraint than we would have if we had no knowledge of the galactic rotation. The reason for this is that the galactic rotation correction is almost parallel to the

solar motion and is therefore unlikely (it means that we have one of the lowest dipole anisotropies of all observers in the Galaxy). However unlikely this state of affairs is *a priori*, it is hard to imagine that the galactic rotation estimate is wildly wrong, and so we must fully incorporate this information. The next step is to subtract the motion of the Galaxy with respect to other galaxies in the vicinity. Here, as we shall see, the situation is not so clear-cut. As before, the *a priori* expectation is that the velocity will decrease. However, as with the galactic rotation we should prepare ourselvles for possible surprises.

We determine the galactic motion with respect to a sample of galaxies around us by means of a 'Tully–Fisher' or 'Faber–Jackson' relation, or some analogous intrinsic correlation between a distance-dependent and a distance-independent quality. These 'L–σ' relations, or analogous relations, are established for a set of galaxies which we believe to be at the same distance from us (for example, in a cluster of galaxies). If we then assume that the correlation found is universal, this allows us to correct redshifts to true distances in a statistical manner by minimizing residuals. The validity of a velocity thereby obtained depends crucially on the assumed universality of the relation. If there are systematic differences in the zero-point of the relations from cluster to cluster, then these will be misinterpreted as peculiar velocities. An effect of this kind must be present at some level. In principle one can test for this by applying the method to clusters at a wide range of distances. Any spurious velocities will increase linearly with distance. These effects, if present at a significant level, do not however present any fundamental limit to the depth of sample for which one can determine the net velocity, as the number of clusters increases sufficiently quickly that the errors remain manageable.

The important question, for our purpose, is whether the source of our peculiar velocity has been localized within the sample. If so, we obviously get a stronger constraint on $\Delta\rho/\rho$ at large scales. It seemed, following the publication by Hart and Davies[66] that this was indeed the case. They studied a sample of galaxies with $v \lesssim 5{,}000$ km s^{-1} using a radio variant of the Tully–Fisher (TF) technique, and found that the net motion of these galaxies with respect to the MBR was very small: $v_{\rm HD} = 130 \pm 70$ km s^{-1}. This result stands in contrast to the earlier result by Rubin *et al.*[67] for a sample of similar depth, although doubts about this result had already been expressed[68,69]. However, the Hart and Davies sample has now been reanalysed[70], and yields a much greater net motion of $\gtrsim 700$ km s^{-1}. An independent sample of similar depth was analysed by de Vaucouleurs and Peters[71], who found an intermediate value, $v_{\rm dV} \approx 300$ km s^{-1}. More recently, Collins *et al.*[72] and Burstein *et al.*[73,74] have found net streaming velocities of ~ 700 km s^{-1} using slightly deeper samples, and are in approximate agreement with the original result of Rubin *et al.*[67]. However, these results had been preceded by those of Aaronson *et al.*[75], who used the infrared TF technique to analyse a sample of galaxies lying in the declination strip visible from Arecibo. They found a net motion of $\lesssim 200$ km s^{-1}, although the geometry of their sky coverage results in increased uncertainty for a velocity vector in the direction favoured by Burstein *et al.*[73,74] and by Collins *et al.*[72]. Another interesting result to emerge from the Aaronson *et al.*[75] sample is that the scatter of individual cluster velocities about the Hubble flow (or of any offsets in the TF relation interpreted as velocities) is very small. They obtained a limit of ~ 200 km s^{-1} on the amplitude of such motions, roughly consistent with the individual cluster motions of ~ 400 km s^{-1} (for a different sample) inferred by Burstein *et al.*[75], and suggesting that the infrared TF technique has very low intrinsic scatter.

The differences between the solutions found, both for net motion and for individual cluster motions, give some measure of the uncertainty associated with these methods. Certainly, the question of whether the source of the galactic motion has been localized remains in doubt. It is probably premature to try to use these data to constrain $\Delta\rho/\rho$ on very large scales. The most reliable datum we have is for the galactic motion of ~ 550 km s^{-1} with respect to the microwave frame, and this results in the line labelled PV in Fig. 2. If a consistent picture for large streaming motions does emerge then this will imply, according to standard theory, sizeable density fluctuations ($\gtrsim 30\%$) on a scale of $\gtrsim 100h^{-1}$ Mpc. Such density fluctuations present a challenge to theorists: for instance, it appears that such fluctuations do not arise naturally in the cold dark matter model. A theory in which non-gaussian fluctuations emerge naturally, such as cosmic strings combined with hot dark matter, offers some prospect of simultaneously allowing large streaming motions and low $\delta T/T$. It should someday be possible to directly measure, or possibly exclude, such density fluctuations with large-scale redshift surveys.

Conclusions

The microwave background isotropy limits, as well as the measured dipole anisotropy and large-scale galaxy distribution, have already succeeded in severely constraining inflationary cosmological models. Indeed, the only survivors require an appeal to rather more contrived models than discussed here. One possibility requires a biasing scheme in which the observed galaxy distribution samples the gravitational effects of a universe with $\Omega = 0.1$, while the true Ω is unity[14,61]. Biasing schemes generally predict that $\sim 90\%$ of the Universe should contain 'failed galaxies', with the baryonic matter forming luminous galaxies only in the rare fluctuation peaks that sample $\sim 10\%$ of the Universe[76,77]. One alternative involves appeal to the vacuum energy density (or cosmological constant Λ): with $\Omega_{\rm vac} = 0.2$, where $\Lambda \equiv \Omega_{\rm vac}$, a combination of cosmological constant and cold dark matter can reduce both $\delta T/T$ and the large-scale peculiar velocity field to acceptable levels, while allowing light to be a good tracer of mass[78]. Another option is to let the cold dark matter (with $\Omega = 1$) decay into relativistic weakly interacting particles at a recent epoch (after galaxies have formed), thereby allowing astronomical measurements today to effectively sample a low-Ω universe[79,80]. This scheme apparently fails to produce flat galactic rotation curves[87] and generates an unacceptable feature in the galaxy correlation function[88], but it is consistent with $\delta T/T$ limits[81]. One could also revive hot dark matter in some non-baryonic guise such as the massive neutrino, with nonlinearity of the hot dark matter at a very recent ($z \approx 1$) epoch, as required by the various constraints such as $\xi(r)$ and the pairwise galaxy peculiar velocity distribution, but now seeding the Universe with large-amplitude small-scale fluctuations (cold dark matter, cosmic strings?) in order to allow some, if not all, galaxies to form at an earlier epoch[7,82,83]. Finally, the wakes produced by cosmic strings could themselves generate large sheet-like structures which might dominate the large-scale galaxy distribution[84–86]. String models naturally yield low (but ultimately observable) $\delta T/T$.

It is of course conceivable that we are completely on the wrong track. Perhaps the primordial fluctuation spectrum was not described by a scale-invariant power law, nor was it gaussian, nor even adiabatic. One might then resort to explosive amplification of primordial seeds to explain galaxy formation and leave little trace of the primordial initial conditions. This still poses the problem, however, of the origin of the large-scale clustering of the galaxies. We do observe large-scale correlations (superclusters and voids): surely these trace initial conditions. If our interpretation of the CBR in the standard Big Bang model is correct, and there is no real alternative, then we inevitably expect to see some residue of these initial conditions in the background radiation anisotropy. It would be too perverse of Nature to have thrown down an impenetrable screen which renders such fluctuations invisible. Eventual detection of $\delta T/T$ on some angular scale is inevitable, and it will surely elucidate one of the most challenging mysteries of the Big Bang theory, namely the origin and the formation of large-scale structure.

1. Guth, A. H. *Phil. Trans. R. Soc.* **A307**, 141-147 (1982).
2. Hawking, S. W. *Phys. Lett.* **115B**, 295-297 (1982).
3. Starobinskii, A. A. *Phys. Lett.* **117B**, 175-178 (1982).
4. Guth, A. & Pi, S.-Y. *Phys. Rev. Lett.* **49**, 1110-1113 (1982).
5. Bardeen, J. M., Steinhardt, P. J. & Turner, M. S. *Phys. Rev.* **D28**, 679-693 (1983).
6. Linde, A. *Phys. Lett.* **129B**, 177-181 (1983).
7. Silk, J. & Turner, M. *Phys. Rev.* **D**. (in the press).
8. Ostriker, J. P. & Cowie, L. *Astrophys. J.* **243**, L127-L132 (1981).
9. Ikeuchi, S. *Publs astr. Soc. Japan* **33**, 211-222 (1981).
10. Zel'dovich, Ya. B. *Mon. Not. R. astr. Soc.* **192**, 663-667 (1980).
11. Vilenkin, A. *Phys. Rep.* **121**, 263-315 (1985).
12. Wyse, R. G. & Jones, B. J. *Astr. Astrophys.* **149**, 144-150 (1985).
13. Blumenthal, G., Faber, S., Primack, J. & Rees, M. *Nature* **311**, 517-525 (1984).
14. Davis, M., Efstathiou, G., Frenck, C. S. & White, S. D. M. *Astrophys. J.* **292**, 371-394 (1985).
15. Frenk, C., White, S. D. M. & Davis, M. *Astrophys. J.* **271**, 417-430 (1983).
16. Efstathiou, G. & Silk, J. *Fundam. Cosmic Phys.* **9**, 1-138 (1983).
17. Bond, J. R. & Efstathiou, G. *Astrophys. J.* **285**, L45-L49 (1984).
18. Vittorio, N. & Silk, J. *Astrophys. J.* **285**, L39-L44 (1984).
19. Peebles, P. J. E. & Groth, E. *Astr. Astrophys.* **53**, 131-140 (1976).
20. Peebles, P. J. E. & Yu, J. T. *Astrophys. J.* **162**, 816-836 (1970).
21. Wilson, M. L. & Silk, J. *Astrophys. J.* **243**, 14-25 (1981).
22. Silk, J. *Nature* **215**, 1155-1156 (1967).
23. Silk, J. *Astrophys. J.* **151**, 459-471 (1968).
24. Sachs, R. K. & Wolfe, A. M. *Astrophys. J.* **147**, 73-90 (1967).
25. Peebles, P. J. E. *Astrophys. J.* L1-L5 (1982).
26. Bond, J. R. & Szalay, A. S. *Astrophys. J.* **274**, 443-468 (1983).
27. Zel'dovich, Ya. B. *Astr. Astrophys.* **5**, 84-89 (1970).
28. Efstathiou, G. & Bond, J. R. *Mon. Not. R. astr. Soc.* **218**, 103-121 (1986).
29. Suto, Y., Sato, K. & Kodama, H. *Astrophys. J.* **292**, L1-L4 (1985).
30. Fixen, D. J., Cheng, E. S. & Wilkinson, D. T. *Phys. Rev. Lett.* **50**, 620-622 (1983).
31. Kaiser, N. *Astrophys. J.* **282**, 374-381 (1984).
32. Hogan, C. *Mon. Not. R. astr. Soc.* **202**, 1101-1126 (1983).
33. Hogan, C. J. & Kaiser, N. *Astrophys. J.* **274**, 7-19 (1983).
34. Vilenkin, A. *Phys. Rep.* **131**, 263-315 (1985).
35. Kaiser, N. & Stebbins, A. *Nature* **310**, 391-393 (1984).
36. Gott, J. R. *Astrophys. J.* **288**, 422-427 (1985).
37. Brandenberger, R. H. & Turok, N. *Phys. Rev.* **D** (in the press).
38. de Lapparent, V., Geller, M. J. & Huchra, J. P. *Astrophys. J.* **302**, L1-L5 (1986).
39. Szalay, A., Bond, J. R. & Silk, J. in *Formation and Evolution of Galaxies and Large Structures in the Universe* (eds Audouze, J. & Tran Thanh Van, J.) 101-128 (Reidel, Dordrecht, 1984).
40. Hogan, C. J. *Astrophys. J.* **284**, L1-L4 (1984).
41. Uson, J. M. & Wilkinson, D. T. *Nature* **312**, 427-429 (1984).
42. Ostriker, J. & Vishniac, E. *Astrophys. J.* **306**, L51-L54 (1986).
43. Penzias, A. A. & Wilson, R. W. *Astrophys. J.* **142**, 419-421 (1965).
44. Melchiorri, F., Melchiorri, B., Ceccarelli, C. & Pietranera, L. *Astrophys. J.* **250**, L1-L4 (1981).
45. Fabbri, R., Guidi, I., Melchiorri, F. & Natale, V. *Phys. Rev. Lett.* **44**, 1563-1566 (1980).
46. Lubin, P., Villela, T., Epstein, G. & Smoot, G. *Astrophys. J.* **298**, L1-L5 (1985).
47. Smoot, G. F., Gorenstein, N. V. & Muller, R. A. *Phys. Rev. Lett.* **39**, 898-901 (1976).
48. Strukov, I. A. & Skulachev, D. P. *Soviet Astr. Lett.* **10**, 3-13 (1984).
49. Mandolesi, N. *et al. Nature* **319**, 751-753 (1986).
50. Fomalont, E. B., Kellermann, K. I. & Wall, J. V. *Astrophys. J.* **277**, L23-L26 (1984).
51. Kellerman, K. I., Fomalont, E. B., Weistrop, D. & Wall, J. *Highlights in Astronomy* (Reidel, Dordrecht, in the press).
52. Lubin, P. M. & Smoot, G. F. *Astrophys. J.* **245**, 1-17 (1981).
53. Lubin, P. M., Melese, P. & Smoot, G. F. *Astrophys. J.* **273**, L51-L54 (1983).
54. Rees, M. J. *Astrophys. J.* **153**, L1-L5 (1986).
55. Negroponte, J. & Silk, J. *Phys. Rev. Lett.* **44**, 1433-1437 (1980).
56. Hogan, C. J., Kaiser, N. & Rees, M. J. *Phil. Trans. R. Soc.* **A307**, 97-109 (1982).
57. Boynton, P. E. & Partridge, R. B. *Astrophys. J.* **181**, 243-253 (1973).
58. Lazenby, A. N. & Davies, R. D. *Mon. Not. R. astr. Soc.* **203**, 1137-1169 (1983).
59. Grishchuk, L. P. & Zel'dovich, Ya. B. *Sov. Astr. Lett.* **22**, 125-129 (1978).
60. Bardeen, J. M. in *Proc. Inner Space-Outer Space Symp.* (eds Kolb, E. W., Turner, M. S., Lindley, D., Olive, K. & Seckel, D.) 212-217 (University of Chicago Press, 1986).
61. Kaiser, N. in *Proc. Inner Space-Outer Space Symp.* (eds Kolb, E. W., Turner, M. S., Lindley, D., Olive, K. & Seckel, D.) 258-263 (University of Chicago Press, 1986).
62. Bardeen, J. M., Bond, J. R., Kaiser, N. & Szalay, A. *Astrophys. J.* **304**, 15-61 (1986).
63. Sunyaev, R. A. & Zel'dovich, Ya. B. *Mon. Not. R. astr. Soc.* **190**, 413-420 (1980).
64. Kaiser, N. *Astrophys. J.* **273**, L17-L20 (1983).
65. Vittorio, N. & Silk, J. *Astrophys. J.* **293**, L1-L4 (1985).
66. Hart, L. & Davies, R. D. *Nature* **297**, 191-196 (1982).
67. Rublin, V. C., Thonnard, N., Ford, W. K. & Robert, M. S. *Astr. J.* **81**, 719-737 (1976).
68. Fall, S. M. & Jones, B. J. T. *Nature* **262**, 547-460 (1976).
69. Schechter, P. L. *Astr. J.* **82**, 569-576 (1977).
70. Stavely-Smith, L. thesis, Univ. Manchester (1985).
71. de Vaucouleurs, G. & Peters, W. L. *Astrophys. J.* **287**, 1-16 (1984).
72. Collins, C. A., Joseph, R. D. & Robertson, N. A. *Nature* **320**, 506-508 (1986).
73. Burstein, D. *et al.* in *Galaxy Distances and Deviations From the Hubble Scale* (eds Madore, B. & Tully, R. B.) (Reidel, Dordrecht, in the press).
74. Lynden-Bell, D. *Q. Jl R. astr. Soc.* **27**, 319-327 (1986).
75. Aaronson, M. *et al. Astrophys. J.* **302**, 536-563 (1986).
76. Silk, J. *Astrophys. J.* **207**, 1-8 (1985).
77. Rees, M. J. *Mon. Not. R. astr. Soc.* **213**, 75P-81P (1985).
78. Peebles, P. J. E. *Astrophys. J.* **284**, 439-444 (1985).
79. Turner, M., Steigman, G. & Krauss, L. *Phys. Rev. Lett.* **52**, 2090-2093 (1984).
80. Gelmini, G., Schramm, D. & Valle, J. *Phys. Lett.* **146B**, 311-317 (1984).
81. Vittorio, N. & Silk, J. *Phys. Rev. Lett.* **54**, 2269-2272 (1985).
82. Achulli, S., Occhionero, F. & Scaramella, R. *Astrophys. J.* **299**, 577-582 (1985).
83. Umemusa, M. & Ikeuchi, S. *Astrophys. J.* **299**, 583-592 (1985).
84. Silk, J. & Vilenkin, A. *Phys. Rev. Lett.* **53**, 1700-1703 (1984).
85. Vachaspati, T. *Phys. Rev. Lett.* **57**, 1655-1657 (1986).
86. Brandenberger, R., Stebbins, A., Turok, N., Veeraraghavan, S. & Silk, J. *Astrophys. J.* (submitted).
87. Flores, R., Blumenthal, G. R., Dekel, A. & Primack, J. R., preprint SCIPP 86/64 (1986).
88. Veeraraghavan, S., Suto, Y. & Silk, J., *Astrophys. J.* (submitted).

ARTICLES

A molecular gradient in early *Drosophila* embryos and its role in specifying the body pattern

Paul M. Macdonald* & Gary Struhl*

Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Ave, Cambridge, Massachusetts 02138, USA

After fertilization, the protein products of the Drosophila *homeobox gene* caudal (cad) *accumulate in a concentration gradient spanning the anteroposterior axis of the developing embryo. Mutations in the* cad gene *that reduce or eliminate the gradient cause abnormal zygotic expression of at least one segmentation gene* (fushi tarazu) *and alter the global body pattern.*

ONE recurring dilemma in developmental biology is the conflict between mosaic models of embryonic patterning which invoke rigid localizations of qualitatively distinct determinants, and regulative models, which postulate morphogen gradients or inductive cascades of cell-cell interactions. In the case of insect development, the classic experiments of Sander[1-4] and others[5-9] have ruled out mosaic models, except possibly for the segregation of the germ line[10], and argued strongly for anteroposterior gradients of morphogens that specify the body pattern. Further support for such a gradient model has come from the analysis of maternal effect mutations, such as *bicaudal*, *oskar* and *bicoid*, which alter the global segment pattern[11-18]. But concrete demonstrations of graded morphogens have been frustratingly absent.

* Present address: Howard Hughes Medical Institute, Center for Neurobiology and Behaviour, Columbia University College of Physicians and Surgeons, 722 West 168 Street, New York, New York 10032.

During the past few years, molecular studies have led to the identification of a conserved structural domain, the homeobox, common to many genes which control segment development[19,20]. In general, homeobox containing genes have tightly restricted spatial roles[21-25]; hence, they are first activated in specific regions of the embryo around the time segments are established[19,26-33]. Recently, we (unpublished observations) and others[34,35] have identified a new homeobox gene, *caudal* (*cad*)[34], which has the unusual property of being expressed during oogenesis and early embryogenesis. Hence, it seemed possible that this atypical gene might have a more global role in organizing the segment pattern. Indeed, circumstantial evidence for such a role has been obtained by Mlodzik *et al.*[34] and Levine *et al.*[35] who observed that *cad* transcripts form an anteroposterior concentration gradient during the syncytial blastoderm stage when the segment pattern is being set up.

We have used antibodies directed against the *cad* protein, together with mutations in the *cad* gene, to examine its develop-

mental role. We find that the *cad* protein is localized in nuclei and is distributed in an anteroposterior gradient throughout the embryo during most of the syncytial blastoderm stage. As observed for the *cad* messenger RNA[34,35], the *cad* protein gradient forms after fertilization, and disappears by the onset of gastrulation, when expression is restricted to a circumferential band of cells near the posterior pole. Further, we show that mutations which appear to inactivate the *cad* gene alter both segmentation and the expression of the pair-rule gene *fushi tarazu* (*ftz*)[36,26,27]. Thus we have directly visualized a graded gene product which acts in specifying the body pattern.

Establishment of a *cad* protein gradient

Rabbit polyclonal antibodies directed against the *cad* portion of a β-galactosidase-*cad* fusion protein (Fig. 1) were used to visualize the distribution of *cad* protein throughout embryogenesis as described (Figs 2 and 3).

During early *Drosophila* development, the single nucleus of the zygote undergoes a series of 13 largely synchronous divisions before cellularization (see refs 37, 38 for detailed descriptions; these divisions define fourteen stages numbered according to the nuclear division which completes each stage). Initially, the proliferating nuclei accumulate in the central, yolky portion of the egg. We detect little if any *cad* protein in the embryo during these early division cycles (stages 1-4); however, beginning in either stage 5 or 6, the cytoplasm and diffuse zones of 'structured cytoplasm'[38,39] around the nuclei stain faintly with the *cad* antibody (data not shown). During stage 7 most of the nuclei begin to migrate towards the periphery; nuclear migration continues during stage 8 and is complete by the end of stage 9 when peripheral nuclei at the posterior pole are incorporated into pole cells. In this period (stages 7-9), *cad* protein accumulates dramatically, especially in the posterior half of the body, forming a smooth concentration gradient along the anteroposterior axis (Fig. 2*a*; see also Fig. 2*b*). In addition, *cad* protein throughout the body becomes progressively more tightly localized to nuclei. The peripheral nuclei then undergo a further 4 rounds of syncytial divisions, during which time (stages 10-13), the concentration gradient persists and spans the length of the embryo (Fig. 2*b*, *c*, *d*). Note that graded nuclear expression of *cad* protein runs throughout the egg, including both peripheral and 'yolk' nuclei (Fig. 2*c*, *d*); posterior pole cell nuclei also stain intensely with the *cad* antibody.

Formation of the *cad* protein gradient does not depend either on nuclear division or zygotic gene expression. Unfertilized eggs collected within 90 minutes of laying show either no *cad* expression, or moderate, relatively uniform expression. In older unfertilized eggs, the levels of *cad* protein rise markedly, especially in the posterior portion of the egg, generating an anteroposterior gradient similar to that normally present in stage 9-13 embryos (data not shown).

Further polarization

Beginning in early stage 14, the pattern of *cad* protein distribution changes dramatically. Initially, antibody labelling subsides progressively from the anterior end of the embryo (Fig. 2*e*), while expression either persists or is enhanced in the posterior half. This results in a transient bipartite staining pattern in which the anterior 40% of the embryo appears devoid of *cad* protein, while relatively high levels of staining remain in the posterior 50%. The progressive loss of *cad* protein then continues into the posterior half of the embryo, extending more quickly along the ventral side (Fig. 2*f*, *g*). At the same time, *cad* protein levels decline in somatic nuclei at the posterior pole (between 0-13% egg length, EL, measured from the posterior pole) although not in the pole cells. These various changes in *cad* expression generate a diffuse posterior ring of *cad* protein which extends more anteriorly on the dorsal than the ventral side. Still later, as the blastoderm cellularizes and gastrulation ensues, this diffuse zone of expression sharpens into a discrete posterior

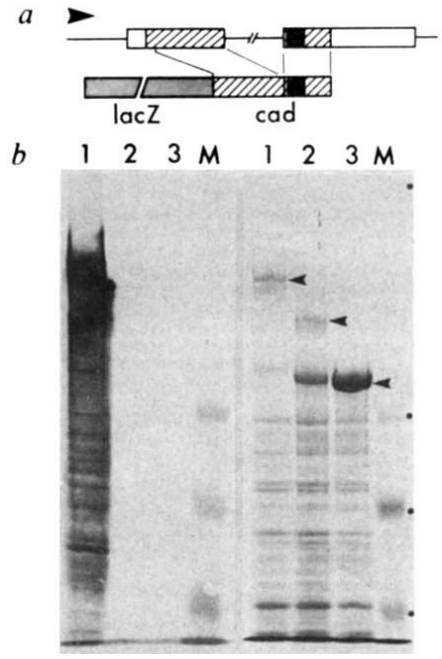


Fig. 1 Antibodies against the *cad* protein. *a*, Structure of the *cad* gene and derivation of a β-galactosidase-*cad* fusion protein. *b*, Specificity of the *cad* antibody. *a*, Zygotic and maternal transcripts of the *cad* gene are composed largely if not exclusively of two exons (wide bars) separated by an ~10 kilobase (kb) intervening sequence. Although the 5′ end of the first exon differs in the various maternal and zygotic transcripts, all have an identical long open reading frame. A 180-base-pair (bp) homeobox is positioned close to the splice junction in the second exon (ref. 34 and P.M.M., unpublished data; open boxes, untranslated 5′ and 3′ portions of the mature maternal transcript; hatched boxes, major open reading frame; black boxes, homeobox domain; arrowhead, direction of transcription). A *lacZ-cad* fusion gene (plc201) was constructed in the vector pUR290 (ref. 64) such that *cad* amino acids 29-425 (a total of 425 is predicted by the DNA sequence) were fused in frame to the carboxy terminus of β-galactosidase (the *cad* gene in plc201 was derived almost entirely from genomic DNA except for a small segment covering the splice junction which was obtained from a cDNA clone; stippled box = the β-galactoasidase coding sequence). *b*, Total protein from IPTG-induced bacteria containing the *lacZ-cad* fusion gene plc201 (lane 1), a *lacZ-even-skipped* fusion gene including the *even-skipped* homeobox domain (lane 2) and the unfused *lacZ* gene from the fusion gene vector (pUR290) (lane 3) was separated by SDS-polyacrylamide gel electrophoresis, electroblotted onto nitrocellulose and probed with affinity-purified rabbit anti-*cad* antibodies prepared as described below (left panel). A duplicate gel stained with Coomassie Blue is shown in the right panel. No cross-reactivity to either β-galactosidase or homeodomain determinants is detectable. The heavy smear extending downward from the *cad* fusion protein bands in lane *C* is due to breakdown products of the unstable fusion protein. Size markers (highlighted by filled circles on the right) have relative molecular masses of 200,000 (M_r 200K) 97.4K, 68K and 43K.

Methods. The *cad* gene was cloned in a systematic screen[29] for new homeobox-containing genes, and the complete coding sequence of the gene determined directly (P.M.M., unpublished data). Positive confirmation that this gene is indeed the *cad* gene described by Mlodzik *et al.*[34] is as follows: (1) *in situ* hybridization experiments show that it is located in the 38E5,6 polytene band interval of salivary gland chromosomes, (2) it generates maternal and zygotic transcripts with sizes and temporal profiles similar to those described[34], (3) the DNA sequence is virtually identical to the partial *cad* cDNA sequence published previously[34]. A *lacZ-cad* fusion gene (plc201) was constructed as described in *a*. After IPTG induction of E. coli strain 71-18 (ref. 64) carrying plc201, the fusion protein was partially purified by preparative SDS-polyacrylamide gel electrophoresis, homogenized in Freunds complete adjuvant and injected subcutaneously at multiple sites in virgin female rabbits. Booster immunizations (in Freunds incomplete adjuvant) were given 2 months later, and serum collected after a further 10 days. Contaminating antibodies directed against β-galactosidase, other *E. coli* proteins, and generic homeodomain determinants were removed by successive passages over affinity columns prepared by coupling the desired ligand to Affigel 10 (Biorad). Ligands used were purified β-galactosidase (BRL), or a crude bacterial extract from cells containing a *lacZ-even-skipped* fusion gene (pleBB1, which includes the *even-skipped* homeobox). A similar column, prepared from the plc201 bacterial extract, was used for affinity purification of the anti-*cad* antibodies. Detailed descriptions of plasmid constructions and antibody purification are available upon request. The Western blot was prepared and stained by standard methods, using an NTB/BCIP alkaline phosphatase detection system (BRL).

ring 3-4 cells wide (Fig. 2*h*; a dorsal stripe of labelling sometimes persists transiently at around 50-60% EL). Note that throughout this process, the changes in *cad* distribution in the peripheral and yolk nuclei are tightly coordinated until the posterior ring is well defined, at which time the yolk nuclei stain poorly or not at all.

Expression of *cad* in subsequent development

At the onset of gastrulation, the presumptive mesoderm, including cells from the most ventral portion of the *cad* posterior ring, invaginate through the ventral furrow; soon thereafter, the pole cells, which continue to label with the *cad* antibody, begin migrating dorsally and anteriorly, surrounded by the remaining cells of the posterior ring. As the germ band extends, the pole cells and possibly some cells from the posterior ring appear to enter the posterior mid-gut invagination (Fig. 2*i*). Subsequently, cells in the developing posterior mid-gut and malpighian tubules (but not in the hindgut) express high levels of *cad* protein. The *cad* staining cells which remain on the outside of the embryo form a segment-length block of cells in the posterior germ band, which ultimately gives rise to parts of the larval anal pads (Fig. 2*j*). Based on double-labelling experiments with a monoclonal antibody directed against *Ultrabithorax* protein products[40], this block of cells appears to lie immediately posterior to parasegment 14 (see ref. 41) and may thus correspond to a cryptic fifteenth parasegment (data not shown).

Three additional features of *cad* expression are notable. First, in favourable preparations of embryos during germ band extension, we observe six narrow stripes of cells positioned at double segment length intervals (Fig. 2*i*). Second, a few bilaterally symmetric pairs of neurons are labelled in the neuromeres of parasegments 1-14 (ref. 42) after germ band shortening (later this labelling is retained in only two to four neuromeres in the thorax and anterior abdomen; Fig. 2*j*). In addition, a small cluster of neuroblasts perhaps belonging to a cryptic fifteenth parasegment labels at the posterior end of the nerve cord (data not shown). Third, persistent *cad* labelling is also observed during the remainder of larval development and in adults in the posterior mid-gut, malpighian tubules, and portions of the genital imaginal disc from which the adult analia may be derived (Fig. 2*j* and data not shown).

Mitotic cycle dependence

The large size and mitotic synchrony of the peripheral nuclei in the syncytial blastoderm stages allow us to monitor the presence and subnuclear distribution of the *cad* antigen throughout the mitotic cycle. As shown in Fig. 3*e*, the *cad* antibody does not label interphase nuclei uniformly, but rather reveals a punctate distribution of *cad* protein which is similar to, although not coincident with, that of AT rich satellite DNA (as revealed by the Hoechst stain which labels DNA in proportion to its AT-content[43]). In contrast, a monoclonal antibody, 8C5, which appears to recognize a ubiquitous chromosomal antigen[70] displays a more uniform and diffuse staining pattern (Fig. 3*f*).

At the onset of mitosis, nuclear *cad* labelling disappears almost completely (as previously noted for *engrailed* protein[44]), although both the Hoechst stain and the 8C5 antibody clearly label condensed chromosomes (Fig. 3*a*-*c*). Because the level of diffuse cytoplasmic staining seems to rise during this period, it is possible that the *cad* protein is actually displaced from condensed chromosomes rather than being rendered inaccessible to the antibody. By late telophase, nuclei begin to label again with the *cad* antibody, although punctuate staining is not evident until interphase.

These observations suggest that *cad* protein may be associated with discrete chromosomal sites during interphase (when transcription occurs) and that these sites may be closely linked or brought close together within the nucleus. Also, they suggest the possibility that *cad* protein may be displaced from the chromosomes during mitosis.

Mutations in *cad* alter segment specification

Four recessive mutations of the *cad* gene have been obtained using EMS mutagenesis ($cad^{1,2,3,4}$; see Fig. 4). Three of these four mutations ($cad^{2,3,4}$), as well as deletions[45] of the *cad* gene region, cause recessive lethality and eliminate detectable levels of *cad* protein in the diffuse posterior ring and later embryonic derivatives where *cad* is normally expressed (for example, Fig. 2*i*, *j*). They therefore behave as apparent null alleles and are referred to subsequently as cad^- mutations.

Homozygous cad^- embryos obtained from heterozygous mothers give rise to almost normal first instar larvae. However, these larvae lack cuticular structures of the posterior terminalia, most notably the anal tuft, parts of the anal pads, and the terminal sense organs[24] (sense organs 1 in ref. 46), all structures supposed to derive from a cryptic tenth abdominal segment[46] (or parasegment 15) (Fig. 4*d*). Thus embryos receiving maternally derived gene product, but lacking a functioning *cad* gene, form a normal segment pattern except for terminal structures derived from cells which display persistent zygotic *cad* expression.

When embryos derived from $cad^-/+$ parents are stained for expression of the *cad* gene product, all embryos before stage 14 show the characteristic patterns of *cad* protein, even though $\frac{1}{4}$ should be homozygous for the cad^- mutation. Thus, normal graded patterns of *cad* protein can be generated from maternally derived products of the *cad* gene, without any zygotic contribution. To examine the consequences of lacking *cad* gene function during this early period of embryogenesis, it is therefore neccessary to eliminate both the maternal and zygotic contributions. This has been done by generating clones of homozygous cad^- germ cells by gamma-ray induced mitotic recombination in $cad^-/+$ larvae (as in ref. 47), and then fertilizing eggs derived from such mutant germ line cells with cad^- sperm (see Fig. 5).

After fertilization of mutant cad^- eggs with sperm from $cad^-/+$ males, two classes of progeny were obtained in approximately equal numbers. The first class give rise to first instar larvae which usually hatched and often developed into normal fertile adults. However, many of these larvae were found to lack portions of the eighth abdominal segment (Fig. 4*c*); less frequently they also showed partial deletions of the fourth abdominal segment, and on occasion, of other, usually even-numbered, abdominal segments. These larvae almost certainly derived from mutant eggs fertilized by wild type sperm, because all of the progeny obtained from mutant eggs fertilized by wild type males developed in this fashion. Thus, even in the absence of maternally derived cad^+ function, zygotic expression from a single, paternally derived cad^+ gene is sufficient to allow normal, or almost normal, development.

Progeny of the second class show a dramatic alteration of the normal body pattern and presumably arise from fertilization of mutant eggs by mutant sperm (Fig. 4e-h). Although the segmentation phenotypes of these larvae are variable both from animal to animal, and between the left and right sides of single animals, certain common features emerge. First, the head and anterior thorax are usually normal, while the remainder of the body is severely shortened owing to variable deletions of many of the posterior segments. In general the second thoracic and odd-numbered abdominal segments are more resistant to deletion, so that most larvae show a phenotype resembling that of alternating segmental deletions caused by mutations in 'pair-rule' segmentation genes[48]. Second, most of the eighth abdominal segment and posterior terminalia (although curiously not some portions of the posterior spiracles) are absent and replaced by small plates of sclerotized cuticle resembling mouthhooks. In addition small clusters of very fine ventral hairs (usually characteristic of the patch of ventral hairs in the middle of the prothorax) lie between these plates and the preceding

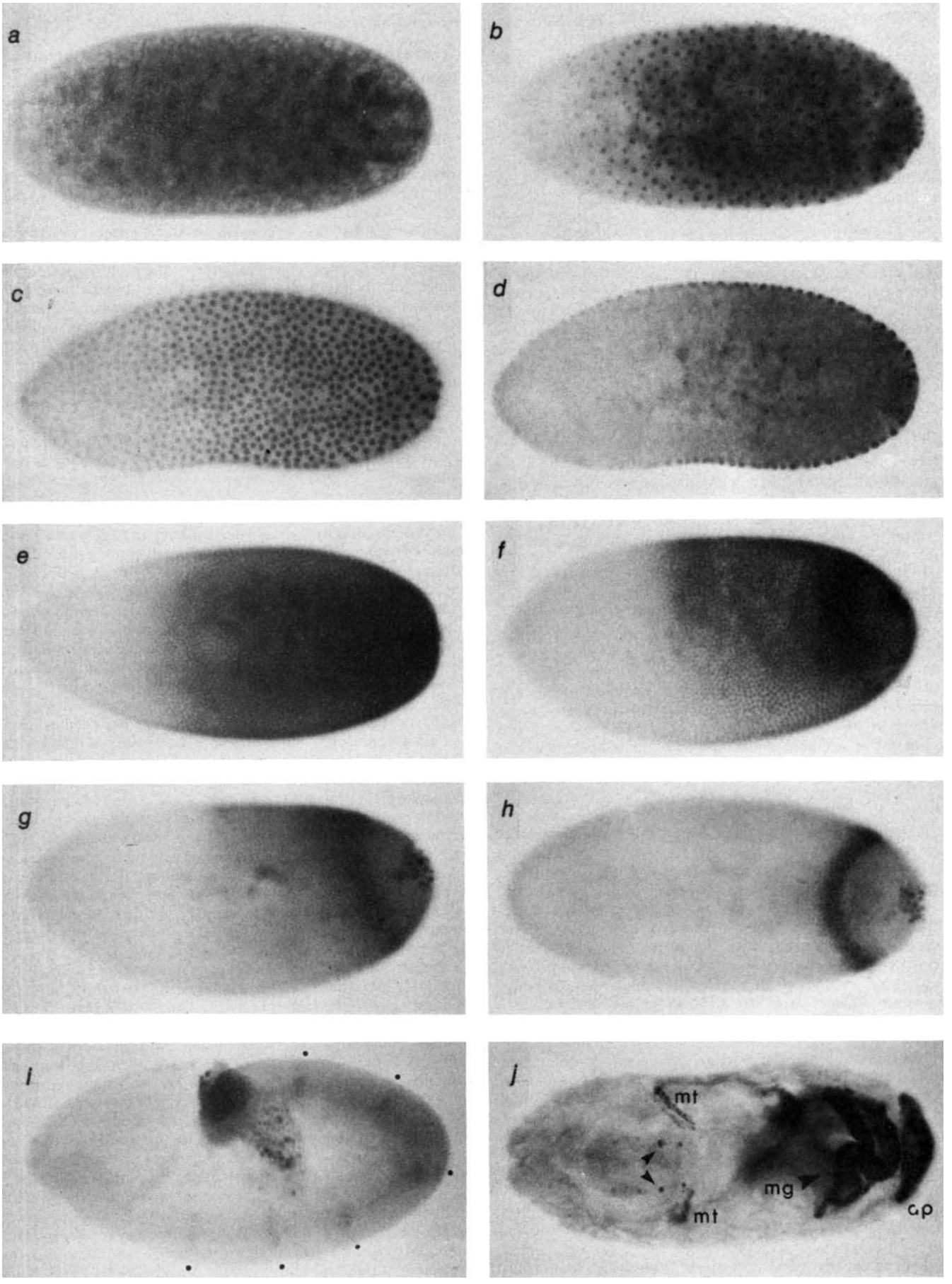


Fig. 2 (legend opposite).

Fig. 2 Immunohistochemical localization of *cad* protein during embryogenesis. All embryos are oriented with anterior to the left and the plane of focus through the middle of the embryo (*a, d, i, j*) or at the periphery (*b, c, e, f, g, h*); embryos before gastrulation are staged according to the nuclear division cycle (for instance, stage 9 falls between the 8th and 9th nuclear divisions). *a*, Stage 8 embryo. All of the nuclei are still located within the central yolky region of the egg. Staining with the *cad* antibody is most intense in the nuclei, but note that islands of cytoplasmic staining surround each nucleus. Graded distribution of the *cad* protein is already apparent, as displayed by the gradual reduction of staining towards the anterior pole. *b*, Stage 10; Nuclear migration to the periphery is complete and the pole cells have budded off. The *cad* protein gradient now extends along the entire body axis, peaking at the posterior pole. This pole to pole concentration gradient is maintained until the beginning of stage 14. Note that it is unlikely to be due to either a fixation artifact or unequal accessibility of the antibody, as both the DNA dye Hoechst 33258 as well as a monoclonal antibody directed against a chromosomal antigen reveal uniform staining throughout this period in triple labelling immunofluorescent experiments (see Fig. 3); the close to uniform distribution of *cad* protein observed in *BicD* embryos (Fig. 5a) also argues against such possibilities. *c, d*, Stage 12 (periphery and inside respectively). The *cad* gradient is confined predominantly to nuclei and is apparent both in peripheral and yolk nuclei. Pole cell nuclei, just visible in *d*, stain as intensely as the neighboring somatic nuclei. *e*, Early stage 14. During stage 14 the distribution of *cad* protein undergoes a dramatic change, becoming progressively more polarized until it is restricted to a single posterior ring 3-4 cells wide at the onset of gastrulation. This embryo shows the early phase of the redistribution in which *cad* protein is removed progressively from the anterior pole. *f*, Early-mid stage 14 (dorsal side at top). Progressive loss of *cad* staining continues extending more quickly along the ventral surface; staining also begins to decline in somatic nuclei between 0-13% egg length (EL; from the posterior pole). Thus, *cad* staining now appears predominantly in a diffuse posterior ring which extends dorsally to about 55% EL. *g*, Mid-stage 14 (dorsal side at top). *cad* staining continues to decline from the anterior and dorsal directions, although intense staining persists in the posterior ring (now somewhat narrower) and in the pole cells. Note that the anterior dorsal boundary of *cad* labelling remains quite distinct, despite the falling level of expression. Note also that here, as in *f*, *cad* expression in the yolk nuclei (out of focus, but visible as dark spots in the center of the embryo) parallels that of the peripheral nuclei, declining progressively in the anterior and ventral portions of the egg. *h*, Late stage 14 (dorsal surface). By the onset of gastrulation, *cad* protein is found in a tightly restricted posterior ring (which is ~3-4 cells wide around the entire embryo) and in the pole cell nuclei. In some preparations, a relatively faint dorsal stripe 1-2 cells wide can be seen anterior to the posterior ring. *i*, Germ band extension (lateral aspect). As the germ band extends dorsally and anteriorly, most *cad* expressing cells from the posterior ring remain at the terminus of the germ band, though some may enter the posterior midgut invagination together with the labelled pole cells. Transient *cad* expression is detected in rows of cells (●) which transect the elongating germ band, spaced at approximately double segment intervals from the block of *cad* staining cells at the end of the germ band. *j*, 15 h embryo (ventral aspect). The *cad* staining can be seen in cells giving rise to portions of the anal pads (*ap*), the posterior midgut (*mg*), and the malpighian tubules (*mt*) which wander in and out of the plane of focus. Staining is also apparent in one or more pairs of bilaterally symmetric nuclei in several neuromeres (arrow heads). Similar neuroblast staining is seen in most if not all of the neuromeres at an earlier stage (after germ band shortening), but is restricted to only a few thoracic or anterior abdominal neuromeres as embryogenesis continues.
Methods. Embryos aged for 0-15 h after fertilization were dechorionated, fixed, devitellinezed, and incubated with the primary antibody as described previously[65,27,44]. The embryos were washed under standard conditions, and incubated with a goat anti-rabbit IgG secondary antibody coupled to biotin (Vector Labs) which had been preadsorbed with 0-12 h embryos. The embryos were then washed again and incubated with a streptavidin-horseradish peroxidase conjugate (BRL). After further washing, the histochemical stain was developed with hydrogen peroxide and diamino-benzidine to yield a permanent signal. Stained embryos were dehydrated in 100% ethanol and mounted directly in a mixture of ~1.5 g ml^{-1} dried canada balsam in methyl salicylate[66]. Preadsorption of the *cad* antibodies with the *cad* affinity resin (see Fig. 1 legend) eliminated the staining patterns shown here.

abdominal segment (usually A7). Note that, despite the severity of the segmental deletions, the overall polarity of cuticular derivatives such as ventral and dorsal hairs is usually unaffected.

Thus, complete absence of the *cad* gene function dramatically alters the normal segment pattern. In parallel with the polarized distributions of *cad* gene product during the syncytial blastoderm stage, anterior head and thoracic segments are usually unaffected, whereas more posterior portions of the body show progressively more severe defects.

Maternal versus zygotic contributions

The extreme segmental defects displayed by cad^- mutant embryos can be rescued either by maternal or zygotic activity alone. Consequently, common features of maternal and zygotic

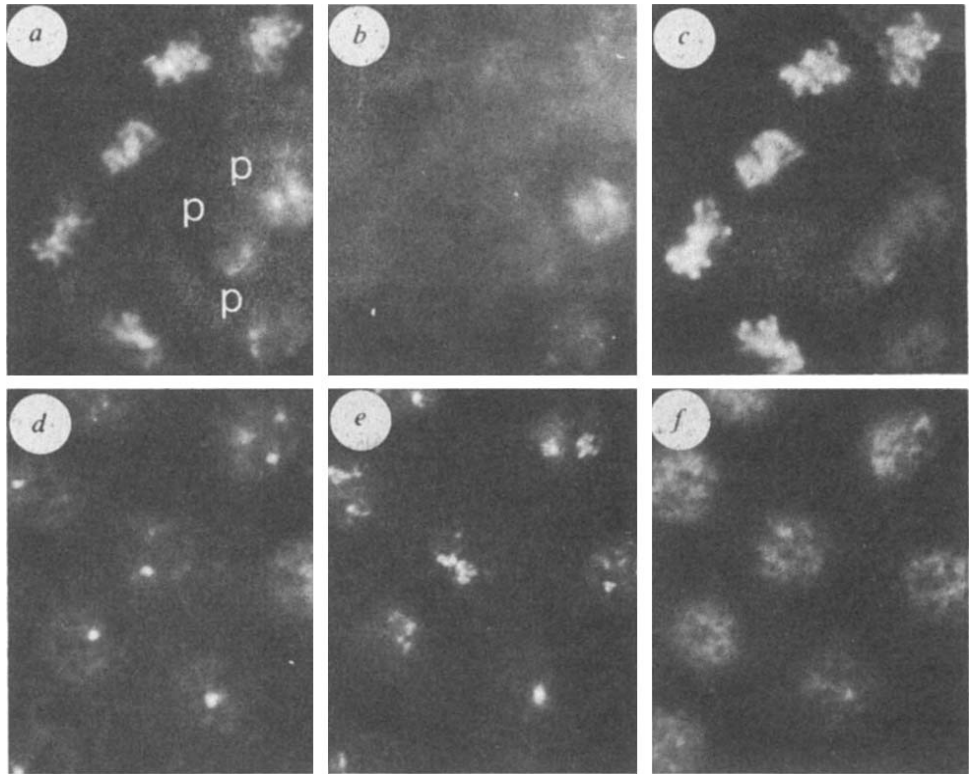


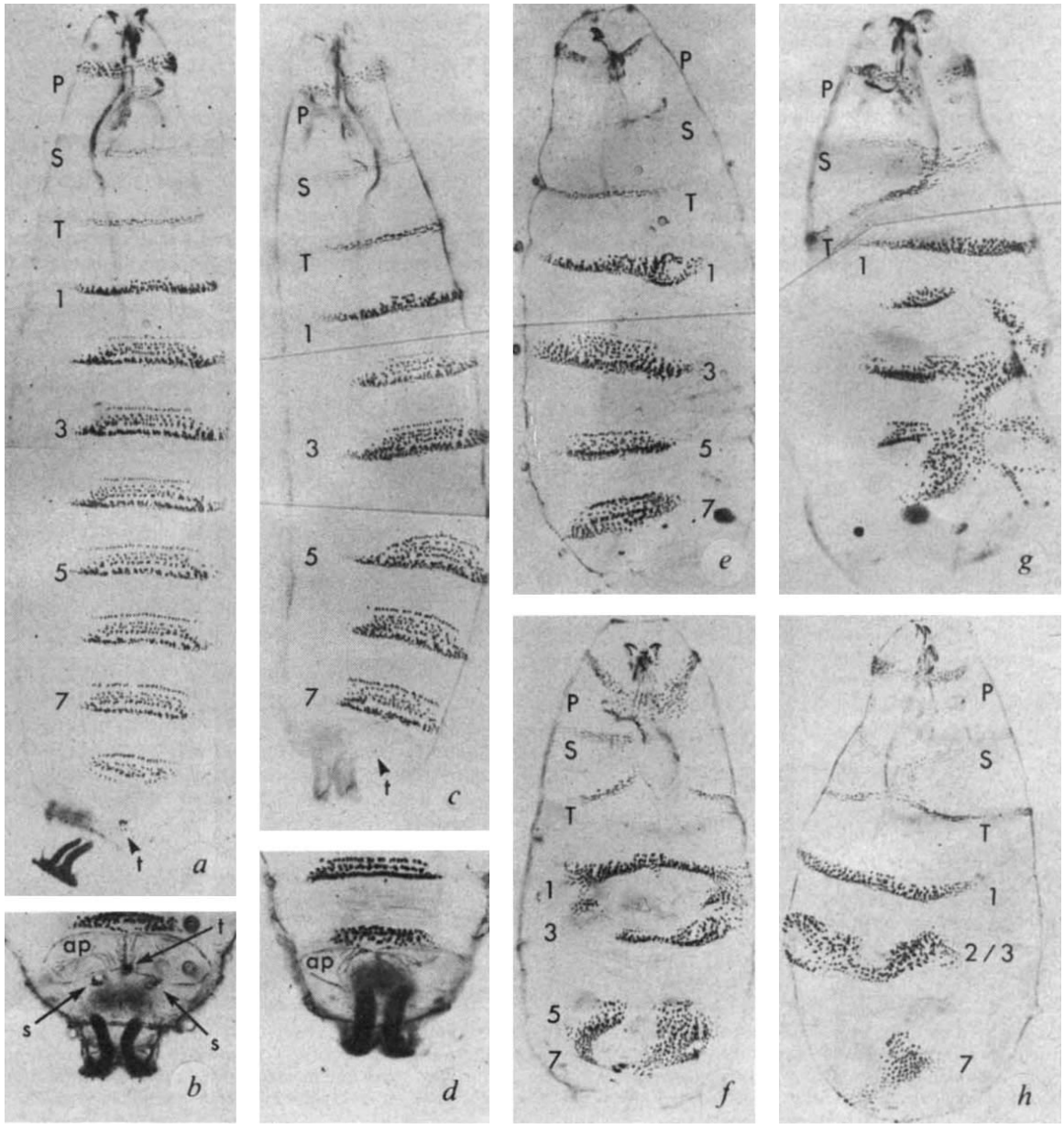
Fig. 3 Mitotic cycle dependence and subnuclear distribution of *cad* staining. The relative distributions of DNA, *cad* protein, and a chromosomal antigen recognized by the monoclonal antibody 8C5 (ref. 70) are shown during the 12th mitotic division (*a, b* and *c*, respectively) and during the interphase preceding this division (*d, e* and *f*). In both bases, the nuclei were located close to the posterior pole (indeed, in *a, b* and *c* the interphase nuclei of three pole cells (*p*) are adjacent to the mitosing syncytial nuclei). Note that the Hoechst stain (*a*) and 8C5 antibody (*c*) clearly show condensed chromosomes in mitosing peripheral nuclei, whereas only faint labelling is detected with the *cad* antibody (*b*). In contrast, interphase peripheral nuclei (*d, e, f*) as well as interphase pole cell nuclei (*a, b, c*) show punctate staining with both the Hoechst stain (*a, d*) and *cad* antibody (*b, e*) and a more diffuse pattern of staining with the 8C5 antibody (*c, f*). Note that the punctate pattern of the Hoechst stain (presumably labelling AT-rich satellite sequences[43]) does not coincide with that revealed by the *cad* antibody.
Methods. Embryos were prepared as in Fig. 2 and probed with the rabbit anti-*cad* antibody as well as the mouse monoclonal antibody 8C5. After washing, the embryos were incubated with two secondary antibodies, a preadsorbed goat anti-rabbit immunoglobulin G (IgG) coupled to biotin (Vector) and a goat anti-mouse IgG coupled to rhodamine (Tago Labs). The embryos were then washed again and incubated with a fluorescein-avidin conjugate (Vector Labs), stained with the Hoechst dye, washed, mounted in 50% glycerol with 2.5 mg ml^{-1} *p*-phenylenediamine added as an anti-bleaching agent, and examined with epi-fluorescence illumination.

protein expression may define critical aspects of *cad* function which are responsible for ensuring normal segmentation. As described above, all embryos derived from $cad^-/+$ embryos show normal graded patterns of *cad* protein until the beginning of stage 14. But after this time $\sim\frac{1}{4}$ fail to show the normal bipartite pattern of expression or the subsequent polarization which gives rise to the posterior ring. Instead, embryos in this novel class display a gradually fading gradient of *cad* protein, although pole cells continue to stain intensely well into gastrulation (Fig. 5*d*; compare with Fig. 2*g*). Paternally rescued embryos ($cad^-/+$ embryos derived by fertilizing cad^- eggs with wild type sperm) resemble maternally rescued cad^- embryos in forming an anteroposterior gradient of *cad* protein, although the levels of *cad* expression are atypically low, close to the limit of detection (data not shown). In stage 14, the level of *cad* antigen rises significantly in posterior portions of paternally rescued embryos, forming the normal bipartite pattern (Fig. 5*e*) and then becoming progressively restricted to the posterior ring (as in Fig. 2*e*-*h*; note that no *cad* protein is expressed in pole cells).

Thus, the common link between the maternal and zygotic contributions to *cad* protein expression appears to be an anteroposterior differential in protein concentration which arises during the syncytial blastoderm stage. This can be generated solely by gradients of maternally derived gene product, or by graded and then more polarized patterns of zygotic expression.

Symmetrical *cad* expression in embryos

If the gradient of *cad* protein plays an instructive role in specifying the body pattern, mutations which cause global changes in the body plan should alter segmentation and the distribution of *cad* protein in a tightly coordinated fashion. Dominant mutations of the *BicD* locus act during oogenesis to alter the axial polarity of the early embryo[15]. Specifically, embryos derived from mutant mothers display a 'double abdomen' phenotype in which the body is composed of two sets of posterior abdominal segments and terminalia arranged in mirror symmetry[15]. When embryos from mothers transheterozygous for two dominant *BicD* mutations (see Fig. 5, legend) are stained with the *cad* antibody, they show corresponding alterations in the normal patterns of *cad* expression. First, in most embryos, *cad* antigen is distributed almost uniformly in nuclei throughout the body until early stage 14 (Fig. 5*a*). The intensity of labelling then



declines in the central portion of the embryo as well as in narrow zones at both ends (Fig. 5*b*). Finally, these embryos display a stable pattern of mirror symmetric posterior rings (Fig. 5*c*). These results show that *BicD* mutations cause a reorganization of the *cad* protein distribution which parallels that of the body pattern.

Expression of *fushi tarazu* in *cad*$^-$ embryos

The tendency of *cad*$^-$ mutant embryos to show pair-rule segmentation defects in the posterior half of the body suggests that one role of early *cad* protein activity may be to coordinate periodic expression of pair-rule genes[48] along the anteroposterior body axis. We therefore examined the expression pattern of the pair-rule gene *fushi tarazu* (*ftz*)[36,26,27] in cad$^-$ embryos using a polyclonal antibody directed against the *ftz* protein (ref. 27). In wild type embryos, the *ftz* protein is expressed in 7 stripes of blastoderm nuclei shortly before the onset of gastrulation: the first six stripes are of equal width (3-4 nuclei) whereas the posterior seventh stripe is wider (5-7 nuclei)[27]. When *cad*$^-$ eggs were fertilized by sperm from *cad*$^-$/+ males, and the resulting embryos stained for *ftz* protein expression at the onset of gastrulation, two classes of progeny were observed in similar numbers. Embryos of the first class almost certainly correspond to paternally rescued *cad*$^-$/+ embryos and approximate wild type *ftz* expression, although stripes 2 and 4 are slightly wider, and stripes 3, 5 and 6 slightly narrower than normal (Fig. 6*a*). In contrast, embryos of the second class almost certainly correspond to homozygous *cad*$^-$ embryos and show highly variable and abnormal patterns of *ftz* expression, stripes 2 and especially 4 being much wider than in wild type embryos, and stripes 3, 5, 6 and 7 being much narrower or absent (Fig. 6*b*). Thus, absence of *cad* protein activity early in embryogenesis causes abnormal periodic expression of the *ftz* protein.

Discussion

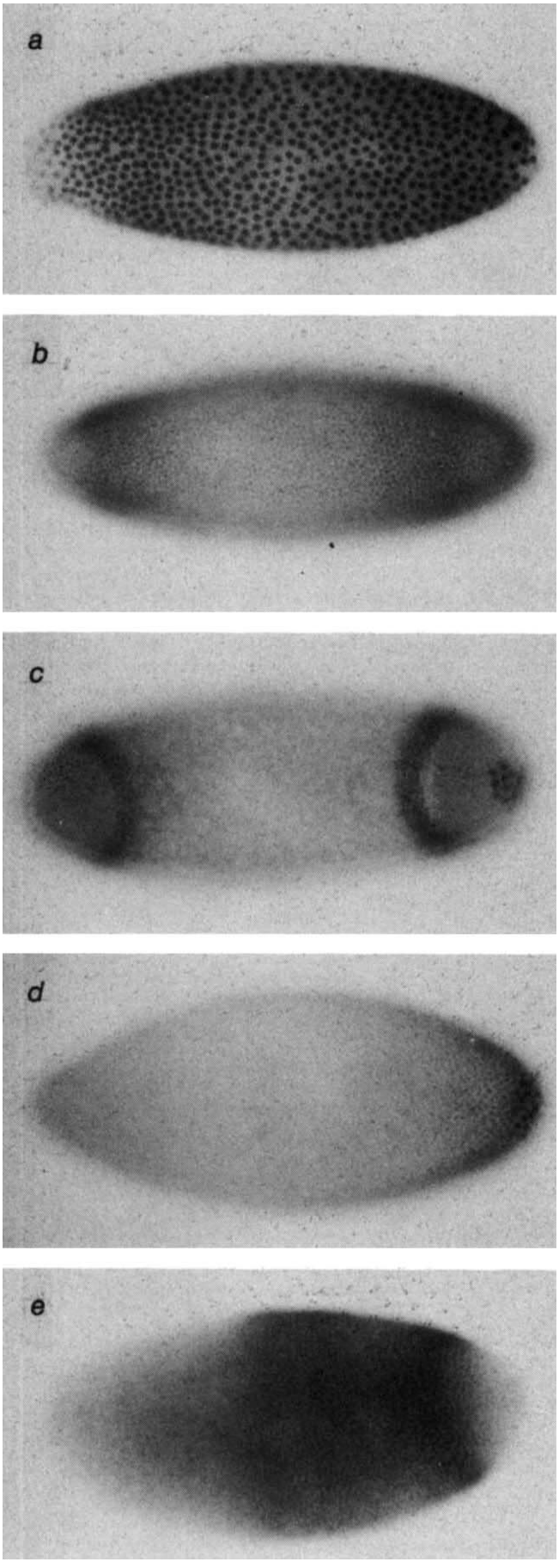
Both embryological[1-9,49-52] and genetic[11-18] experiments provide compelling evidence that the segment pattern of the insect body is specified by a dynamic system of spatial cues generated early in development. Moreover, they have led to explicit proposals that this system consists of one or more morphogen gradients organized by the spreading influences of determinants at the anterior and posterior poles[2,17,18,53]. Here, we visualize a protein gradient which arises during the period when the body plan is established. Moreover, we show that the formation of this gradient depends on maternal factors such as the *BicD* gene[15] which influence the localization of anterior and posterior polar determinants[17,18]. Finally, we demonstrate that activity of the gradient protein is critical for setting up the normal segment pattern. Thus, we believe we have identified a graded morphogen which responds to early spatial cues and participates in organizing the body plan.

Role of the *cad* protein gradient

The developing pattern of *cad* protein expression is complex, involving first the accumulation of *cad* protein in the form of a continuous anteroposterior gradient, followed by progressive loss of the protein from the anterior pole until it is restricted to a ring of cells close to the posterior pole. Hence, the morphogenic role of the *cad* gene product could depend on the concentration gradient *per se*, or more simply, on the consequent generation of an anteroposterior differential of *cad* protein activity. Two lines of evidence support the view that graded *cad* expression *per se* is required for normal segmentation. First, embryos lacking the *cad* function show a spectrum of graded segmental defects which appear to correlate with the graded expression of *cad* protein during the syncytial blastoderm stage. Hence, the requirement for *cad* function appears proportional to the relative levels of protein along the body. Second, early graded expression of maternally derived gene product is sufficient to rescue the development of *cad*$^-$ zygotes (except for derivatives of the posterior terminalia which require persistent zygotic expression). Neither line of evidence is compelling, however. Indeed, we also observe that zygotic expression, which generates first an abnormally shallow gradient, and then a more normal bipartite distribution can largely rescue the *cad*$^-$ mutant phenotype. This last result argues against attaching a critical significance to the absolute concentration of *cad* protein along

Fig. 4 Segmental phenotypes resulting from loss of *cad* gene function. Four recessive mutations of the *cad* gene (*cad*^{1,2,3,4}) were obtained after EMS mutagenesis and the phenotypes of embryos lacking either maternal gene function (*c*), zygotic gene function (*d*), or both (*e–h*) have been determined as described below. Micrographs show the ventral cuticle patterns of first instar larvae (dark field illumination); the ventral dentical bands associated with the pro- (*P*), meso- (*S*) and metathorax (*T*) and the odd numbered abdominal segments (1, 3, 5, 7) are labelled; t, anal tuft (a patch of ventral denticals positioned just posterior to the anal pads); *a*, wild type; *b*, wild type posterior terminalia. Note the bilaterally symmetric anal pads (ap), the anal tuft (t) and the terminal sense organs (s) which lie just posterior and to either side of the tuft. *c*, paternally rescued *cad*^{2}/+ larva derived from a homozygous *cad*^{2} oocyte. Note the absence of the dentical bands associated with abdominal segments 4 and 8; neither segment is entirely deleted, as dorsal structures from both segments (not visible here) remain. *d*, Posterior portion of a homozygous *cad*^{3} larva derived from *cad*^{3}/+ parents. Larvae of this genotype are generally normal except for the absence of the anal tuft and portions of the anal pads. The terminal sense organs also appear to be absent or malformed. *e–h*, *cad*^{2}/*cad*^{3} larvae derived from homozygous *cad*^{2} oocytes. In general, embryos of this genotype usually form normal heads, part or all of the prothorax, and then a spectrum of progressively more extreme segmental deletions along the anteroposterior axis. As described in the text, an 'even-skipped' pair-rule modulation is often apparent (as in *e* and *f*), though more extreme deletions (*h*) or left-right differences in segmental deletions (*g*) can obscure this aspect of the phenotype. The polarity of cuticular structures within each segment is usually normal, though local perturbations are sometimes present in embryos such as *g*.

Methods. The *cad* gene maps to the polytene band interval 38E5,6 as determined by *in situ* hybridization experiments (ref. 34 and our unpublished data), a region which is deleted by a number of well characterized chromosomal deficiencies[45,67]. Two of these deficiencies, *Df(2L)TW1* and *Df(2L)TW2*, overlap, deleting a common region containing polytene bands 38A7-B1–38E6-E9 (ref. 45); a third deficiency, *Df(2L)Sd*^{R68} (37A/B-38E2-E5; ref. 67), deletes much of this common region, but not the 38E5,6 interval (we have confirmed these chromosomal assignments by determining that the *Df(2L)Sd*^{R68} chromosome, but not the *Df(2L)TW1* or *Df(2L)TW2* chromosomes can generate *cad* gene transcripts (P.M.M., unpublished data). We therefore performed a standard F2 screen of ~4,000 chromosomes for EMS-induced mutations which cause partial or complete lethality when *in trans* to *Df(2L)TW2*. All mutations were then retested against the remaining two deletions to eliminate most of the non-*cad* mutations. Of the remaining 30 mutations, 4 fall into a single complementation group which appears to define the *cad* gene by the following criteria: (1) all four mutations specifically cause loss of larval or adult structures which derive from embryonic or larval cells showing persistent zygotic expression of the *cad* protein; (2) approximately $\frac{1}{4}$ of the embryos generated by flies heterozygous for the *cad*^{2}, *cad*^{3} or *cad*^{4} mutations show no *cad* protein attributable to zygotic expression of the *cad* gene (as is observed for progeny from *Df(2L)TW1*/+ and *Df(2L)TW2*/+ parents), and (3) *cad*^{2}/+ embryos derived from *cad*^{2}/*cad*^{2} oocytes express *cad* protein during the syncytial blastoderm stage and in all subsequent stages, whereas *cad*^{2}/*cad*^{3} embryos derived from *cad*^{2}/*cad*^{2} oocytes do not express *cad* antigen at any time (*cad*^{1} homo- or hemizygotes express *cad* antigen, but are also able to survive to adulthood, generating male and female flies lacking anal plates; consequently, this mutation is likely to reduce rather than eliminate *cad* gene function). Finally note that although the three apparent null mutations of *cad* eliminate detectable *cad* protein expression, they may nevertheless retain low levels of *cad* protein activity and hence not be true amorphic mutations. Germ line clones of homozygous *cad*^{2} and *cad*^{3} cells were generated by standard means (see, for example ref. 47) using a dominant female sterile mutation, *Fs(2)1* (courtesy of Janos Szabad and Trudi Schupbach). Briefly, *b pr cad*^{2}/*Fs(2)1* (or *b pr cad*^{3}/*Fs(2)1*) larvae were γ-irradiated during the first instar and the resulting adult females crossed to *b pr cad*^{3}/*CyO, pr* (or *b pr cad*^{2}/*CyO, pr*) males. *Fs(2)1*/+ female germ cells either give rise to woefully abnormal eggs or no eggs at all; however, homozygous *Fs(2)1*$^+$ germ cells resulting from mitotic recombination proximal to the *Fs(2)1* mutation form normal eggs. Individual females laying normal eggs were selected and their progeny examined. Most (~90%) of these females generated the two classes of progeny described in the text, indicating that most recombination events that occur proximal to the *Fs(2)1* gene also occur proximal to the *cad* gene. Moreover, the paternally rescued (*b pr cad*$^-$/*CyO, pr*) progeny from these females often survived to adulthood and were invariably purple in phenotype, confirming that they resulted from proximal mitotic recombination events (the remaining few females laying normal eggs generally gave rise to curly-winged progeny which had phentoypically wild type or purple eyes, indicating that mitotic recombination had occurred proximal to the *Fs(2)1* gene, but distal to the tightly linked *cad* and *pr* genes). Mutant larvae were dechorionated and dissected, when necessary, and mounted directly in a 1:1 mixture of lactic acid and Hoyer's mountant[68,69].



the body, and suggests that a coarse anteroposterior differential in *cad* function is required for normal segmentation.

Irrespective of which component of early *cad* expression is responsible for its morphogenic function, it is plausible that *cad* serves as a molecular link between more primary spatial determinants and the localized expression of zygotic control functions. For example, graded or differential expression of *cad* protein might be responsible for regulating the spatial expression of particular gap genes such as *hunchback*, *Kruppel* and *knirps*[46] which would then establish the segment pattern via the correct activation of pair-rule and homeotic genes[54-56]. Alternatively, graded *cad* expression might directly influence pair-rule and homeotic gene functions, some of which normally display anteroposterior gradients in their spatial expression[28,29,57,58]. Our demonstration that cad^- embryos alter both segmentation and the periodic expression of the pair-rule gene *ftz* over a large region of the body provide direct support for the view that *cad* protein regulates the expression of pair-rule and homeotic genes either directly or via its effects on particular gap genes. The presence of a homeobox domain within the *cad* protein and our finding that *cad* protein is preferentially localized in transcriptionally active interphase nuclei both support this interpretation.

Generation of the *cad* protein gradient

If, as we propose, the dynamic pattern of *cad* expression is a response to the influence of more primary polar determinants, the mechanism by which the *cad* protein gradient forms is likely to be of considerable interest. One striking feature of our results is that *cad* protein expression appears to be graded both in peripheral and yolk nuclei and shifts posteriorly in these two populations of nuclei in a tightly coordinated fashion. Thus, as suggested previously by the initial pattern of activation of the segmentation gene *Kruppel*[59], the generation and interpretation of graded spatial information in insect embryos need not be a special attribute of the peripheral egg cortex, but may occur uniformly throughout the egg cell.

Fig. 5 Distribution of *cad* protein in *BicD* and cad^- mutants. All embryos are oriented with anterior to the left and the plane of focus at the periphery of the egg. *a*, Stage 12 *BicD* embryo; *cad* protein is most abundant near the posterior pole, declines slightly towards the middle of the embryo and remains at a constant high level into the anterior region. The light staining at the anterior tip is commonly observed. Some variation in the overall pattern is seen in a population of embryos, as expected from the varying segmental phenotypes seen in populations of mutant embryos. *b*, Mid-stage 14 *BicD* embryo. Expression in the posterior ring, including the refinement of the initially broad region of staining, occurs roughly as normal, except that the ring forms at both ends. The pole cell nuclei (which are not in this plane of focus, but see *c*) stain intensely, as normal; there are no pole cells (or equivalent staining) at the anterior pole. *c*, Late stage 14 *BicD* embryo (dorsal aspect). Posterior and anterior rings are fully formed, both 3-4 nuclei wide. The width of the anterior ring can vary considerably, sometimes forming a stripe 10 or more cells wide. *d*, Mid-stage 14 cad^- embryo derived from a $cad^-/+$ oocyte. The last traces of the gradient of *cad* protein can be seen near the labelled pole cell nuclei (see Fig. 2*g* for a wild type embryo at a comparable stage). *e*, Early-mid stage 14 embryo derived from a cad^2 oocyte fertilized by wild type sperm (dorsal aspect). *cad* protein expression is similar to a comparably staged wild type embryo (Fig. 2*f*) except that no staining is observed in the pole cells. **Methods.** Antibody staining was as described (Fig. 2, legend), except for *e*; in this case, a strepavidin-alkaline phosphatase conjugate was used in place of the strepavidin-horseradish peroxidase conjugate, and the histochemical stain generated with the *black* kit (Vector Labs). *BicD* embryos were obtained from mothers transheterozygous for two different *BicD* alleles, $BicD^{71.34}$ and $BicD^{IIIF.48}$ (ref. 15). Embryos that were cad^- were obtained as $\frac{1}{4}$ of the progeny of a cross between mothers heterozygous for *Df(2L)TW2* and fathers heterozygous for *Df(2L)TW1*; both deficiencies delete the *cad* gene[45]. Similar results were obtained using embryos from selfcrosses of balanced cad^{2-4} mutants. The cad^2 eggs fertilized by wild type sperm were obtained by fertilizing females carrying germ line clones of the cad^2 mutation with cad^+ males (see Fig. 4).

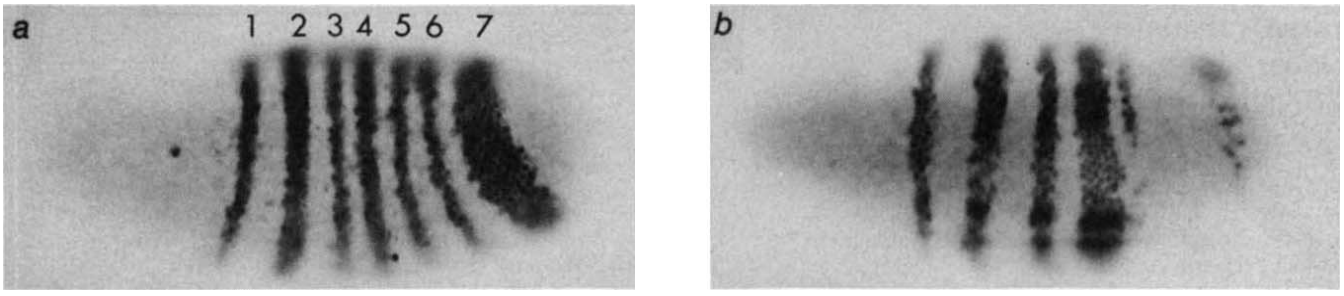


Fig. 6 Altered *ftz* protein expression in paternally rescued and cad^- mutant embryos; $cad^2/+$ (*a*) and cad^2/cad^3 (*b*) embryos derived from homozygous cad^2 oocytes, stained for *ftz* protein expression at the onset of gastrulation (lateral aspect, anterior to the left, dorsal at the top). Wild type embryos at this stage normally express six stripes of *ftz* protein of equal width followed by a broader seventh stripe. In paternally rescued $cad^-/+$ embryos (*a*), stripes 2 and 4 are slightly wider, and stripes 3, 5 and 6 are slightly narrower than in wild type. This tendancy is dramatically accentuated in cad^- embryos (*b*), leading to the loss of stripe 6 and severe reduction of stripes 5 and 7. Immunohistochemical staining as in Fig. 4*e*, except that the NBT/BCIP detection system (BRL) was used instead of the Vector Labs *black* kit.

Although uniformly high levels of *cad* RNA are found throughout the early embryo[34,35], we detect *cad* protein only after the fourth or fifth nuclear division. Soon thereafter, as the *cad* protein gradient forms, the distribution of *cad* RNA also becomes graded. This correlation suggests that a single process is responsible for localizing both the mRNA and protein. Because graded *cad* protein expression can be generated in unfertilized eggs as well as in cad^- embryos derived from heterozygous mothers, this process must depend on post-transcriptional events, rather than on graded zygotic transcription.

In principle, formation of the *cad* protein gradient might reflect a passive response to changes in the distribution of the RNA (for example, mobilization of the transcripts, or local degradation by a specific, spatially restricted nuclease). Alternatively, both the RNA and protein gradients might be controlled at the translational level. One example of how this could occur is through differential initiation of translation along the anteroposterior body axis. Assuming that mRNA associated with ribosomes is shielded from nonspecific nucleases[60,61], efficiently initiated translation near the posterior pole would produce a high level of *cad* protein and at the same time stabilize the RNA, while infrequently translated *cad* RNA at the anterior pole would be prone to degradation. Analysis of the *cad* gene (P.M.M., unpublished data) provides suggestive evidence that supports a translational control model, as the various maternal and zygotic transcripts include sets of in-frame start and stop codons upstream from the probable initiator AUG codon; similar elements clearly act in translational control of the yeast GCN4 mRNA[62,63]. Polar determinants, such as those posited by Sander[2] and others[9,17,18], might regulate translation of *cad* mRNA either directly, or indirectly through their effects on other control molecules.

We thank Barry Ganetzky, Trudi Schupbach, Janos Szabad, Eric Wieschaus, and Ted Wright for fly stocks, Seymour Benzer for the monoclonal antibody 8C5, Sean Carroll and Matthew Scott for the *ftz* antibody, Sean Carroll, George Deegan, Tad Goto, and Ariel Ruiz i Altaba for advice or assistance in preparation of antibodies, Anne Ephrussi, Janice Fischer, Peter Lawrence, Christianne Nüsslein-Volhard and Robin Wharton for critical comments on the manuscript, and Tom Maniatis for advice and support. This work was supported by the U.S. National Institute of Health (Program Project Grant 1PO1GM29301 to Tom Maniatis), the Howard Hughes Medical Institute, and the Damon Runyon-Walter Winchell Cancer Fund (Postdoctoral fellowship DRG-768 to P.M.M.).

Received 22 September; accepted 10 November 1986.

1. Sander, K. *Wilhelm Roux Arch. dev. Biol.* **151**, 430-497 (1959).
2. Sander, K. *Wilhelm Roux Arch. dev. Biol.* **151**, 660-707 (1960).
3. Sander, K. in *Symposium on the Germ Cells and Earliest Stages of Development* (ed. Pallanza, M.) 338-353 (Institut International d'Embryololgie, 1961).
4. Herth, W. & Sander, K. *Wilhelm Roux Arch. dev. Biol.* **172**, 1-27 (1973).
5. Sander, K. *Adv. Insect Physiol.* **12**, 125-238 (1976).
6. Schubiger, G. & Wood, W. J. *Am. Zool.* **17**, 565-576 (1977).
7. Schubiger, G., Moseley, R. C. & Wood, W. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2050-2053 (1977).
8. Vogel, O. *Wilhelm Roux Arch. dev. Biol.* **182**, 9-32 (1977).
9. Frohnhöfer, H. G., Lehmann, R. & Nusslein-Volhard, C. *J. Embryol. exp. Morphol.* (in the press).
10. Illmensee, K. & Mahowald, A. P. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1016-1020 (1974).
11. Nüsslein-Volhard, C. *Wilhelm Roux Arch. dev. Biol.* **183**, 249-268 (1977).
12. Nüsslein-Volhard, C. in *Determinants of Spatial Organisation* (eds Subtelney, S. & Konigsberg, I. R.) 185-211 (Academic, New York, 1979).
13. Bull, A. *J. exp. Zool.* **161**, 221-242 (1966).
14. Lohs-Schardin, M. *Wilhelm Roux Arch. dev. Biol.* **191**, 28-36 (1982).
15. Mohler, J. & Wieschaus, E. F. *Genetics* **112**, 803-822 (1986).
16. Schüpbach, T. & Wieschaus, E. *Wilhelm Roux Arch. dev. Biol.* **195**, 302-317 (1986).
17. Lehman, R. & Nüsslein-Volhard, C. *Cell* **47**, 141-152 (1986).
18. Frohnhöfer, H. G. & Nüsslein-Volhard, C. *Nature* **324**, 120-126 (1986).
19. McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. & Gehring, W. J. *Nature* **308**, 428-433 (1984).
20. Scott, M. P. & Weiner, A. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4115-4119 (1984).
21. Lewis, E. B. *Nature* **276**, 565-570 (1978).
22. Duncan, I. & Lewis, E. B. in *Developmental Order: Its Origin and Regulation, 40th Symposium of the Society for Developmental Biology* (ed. Subtelny, S.) 533-554 (Liss, New York, 1982).
23. Struhl, G. *Nature* **293**, 36-41 (1981).
24. Struhl, G. *J. Embryol. exp. Morphol.* **76**, 297-331 (1983).
25. Struhl, G. *Nature* **318**, 677-680 (1985).
26. Hafen, E., Kuroiwa, A. & Gehring, W. J. *Cell* **37**, 833-841 (1984).
27. Carroll, S. B. & Scott, M. P. *Cell* **43**, 47-57 (1985).
28. Harding, K., Rushlow, C., Doyle, H. J., Hoey, T. & Levine, M. *Science* **233**, 953-959 (1986).
29. Macdonald, P. M., Ingham, P. & Struhl, G. *Cell* (in the press).
30. Akam, M. *EMBO J.* **2**, 2075-2084 (1983).
31. Akam, M. E. & Martinez-Arias, A. *EMBO J.* **4**, 1689-1700 (1985).
32. Levine, M., Hafen, E., Garber, R. L. & Gehring, W. J. *EMBO J.* **2**, 2037-2046 (1983).
33. Harding, K., Wedeen, C., McGinnis, W. & Levine, M. *Science* **229**, 1236-1242 (1985).
34. Mlodzik, M., Fjose, A. & Gehring, W. J. *EMBO J.* **4**, 2961-2969 (1985).
35. Levine, M. *et al. Cold Spring Harb. Symp. Quant. Biol.* **50**, 209-222 (1985).
36. Wakimoto, B. T. & Kaufman, T. C. *Devl Biol.* **81**, 51-64 (1981).
37. Zalokar, M. & Erk, I. *J. Microsc. Biol. Cell.* **25**, 97-106 (1976).
38. Foe, V. E. & Alberts, B. M. *J. Cell Sci.* **61**, 31-70 (1983).
39. Karr, T. L. & Alberts, B. M. *J. Cell Biol.* **102**, 1494-1509 (1986).
40. White, F. A. H. & Wilcox, M. *Cell* **39**, 163-171 (1984).
41. Martinez-Arias, A. & Lawrence, P. A. *Nature* **313**, 639-642 (1985).
42. Hartenstein, V. & Campos-Ortega, J. A. *Wilhelm Roux's Arch. dev. Biol.* **193**, 308-325 (1984).
43. Weisblum, B. & Haenssler, E. *Chromosoma* **46**, 255-260 (1974).
44. DiNardo, S., Kuner, J. M., Theis, J. & O'Farrell, P. H. *Cell* **43**, 59-69 (1985).
45. Wright, T. R. F., Hodgetts, R. B. & Sherald, A. F. *Genetics* **84**, 267-285 (1976).
46. Sato, T. & Denell, R. E. *Devl Biol.* **116**, 78-91 (1986).
47. Wieschaus, E. & Szabad, J. *Devl Biol.* **68**, 29-46 (1979).
48. Nüsslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795-801 (1980).
49. Yajima, H. *J. Embryol. exp. Morphol.* **8**, 198-215 (1960).
50. Yajima, H. *J. Embryol. exp. Morphol.* **12**, 89-100 (1964).
51. Kalthoff, K. *Wilhelm Roux Arch. dev. Biol.* **168**, 63-84 (1971).
52. Kalthoff, K. *Photochem. Photobiol.* **18**, 355-364 (1973).
53. Meinhardt, H. *Models for Biological Pattern Formation* (Academic, London, 1982).
54. Ingham, P., Ish-Horowicz, D. & Howard, K. *EMBO J.* **5**, 1659-1665 (1986).
55. Carroll, S. B. & Scott, M. P. *Cell* **45**, 113-126 (1986).
56. Meinhardt, H. *J. Cell Sci.* (1986).
57. Kilchherr, F., Baumgartner, S., Bopp, D., Frei, E. & Noll, M. *Nature* **321**, 493-499 (1986).
58. Weir, M. P. & Kornberg, T. *Nature* **318**, 433-439 (1985).
59. Kniple, D. C., Seifert, E., Rosenberg, U. B., Preiss, A. & Jackle, H. *Nature* **317**, 40-44 (1985).
60. Zitomer, R. S., Montgomery, D. L., Nichols, D. L. & Hall, B. D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3627-3631.
61. Losson, R. & Lacroute, F. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5134-5137 (1979).
62. Mueller, P. P. & Hinnebusch, A. G. *Cell* **45**, 201-207 (1986).
63. Tzamarias, D., Alexandraki, D. and Thireos, G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4849-4853 (1986).
64. Ruther, U. & Muller-Hill, B. *EMBO J.* **2**, 1791-1794 (1983).
65. Mitchison, T. J. & Sedat, J. *Devl Biol.* **99**, 261-264 (1983).
66. Struhl, G. *Devl Biol.* **84**, 372-385 (1981).
67. Brittnacher, J. G. & Ganetzky, B. *Genetics* **103**, 659-673 (1983).
68. Van der Meer, J. M. *Drosophila Inf. Serv.* **52**, 160.
69. Struhl, G. *Nature* **308**, 454-457 (1984).
70. Fujita, S. C., Zipursky, S. L., Benzer, S., Ferrus, A. & Shotwell, S. L. *Proc. natn Acad. Sci. U.S.A.* **79**, 7929-7933 (1982).

Infrared variability of the BL lacertae object OJ287 since its outburst in 1983

W. K. Gear, E. I. Robson & L. M. J. Brown

School of Physics and Astronomy, Lancashire Polytechnic, Preston PR1 2TQ, UK

In early 1983, OJ287 was seen to undergo an outburst in its optical and infrared emission[1]. Since then optical monitoring studies[2–4] have shown a general decline in the source brightness, with considerable fluctuations. We have monitored the near-infrared emission since the outburst. The fluxes have fluctuated considerably, with the lowest recordings being an order of magnitude less than those measured during the outburst. We have found an excellent correlation between infrared flux and spectral index; as the source gets fainter the spectrum gets steeper, and vice versa. We interpret this in terms of outbursts being due to injection or reacceleration of electrons with a constant energy index which subsequently steepens as the electrons are affected by radiation losses.

We began observing OJ287 in February 1983 as part of a larger programme studying the infrared to millimetre variability properties of a sample of Blazars[5,6]. Observations have been made at J (1.25 μm), H (1.65 μm), K (2.20 μm) and L (3.45 μm) on 16 different nights over the period February 1983 to April 1986. All data were obtained from the United Kingdom Infrared Telescope (UKIRT), using common-user InSb photometers. Calibration was performed against stars from the observatory standards list.

The results of the observations and calibration procedure are shown in Table 1. In Table 2, we also list the infrared spectral indices, obtained by a least-squares fit method, for the 14 of these occasions on which we had complete JHKL coverage. We also checked each spectrum for departures from a power law. Given the rather small wavelength range of our data, and taking into account the uncertainties, we found no significant departure from a single power law at any epoch. Figure 1 plots the variability of the infrared emission over the period 1983–86, where to avoid overcrowding, we have averaged together spectra taken over a period of <1 week. Note that none of the spectra in Fig. 1 has been displaced; the different fluxes represent genuine variability. The extent of variability shown in Fig. 1 is such that to be sure of defining the overall continuum of OJ287, one would need to obtain the whole spectrum over no more than a single night, and preferably an even shorter time than this.

Table 1 Infrared observations of OJ287

	J	H	K	L
3 February 1983	31.0±0.5	42.2±0.6	59.4±1.2	101.6±3.6
5 February 1983	34.2±0.4	47.6±0.3	64.2±0.3	101.6±3.3
6 March 1983	38.5±1.3	51.4±3.4	69.0±3.9	119.0±10.0
29 January 1984	14.7±0.2	22.0±0.2	29.6±0.2	53.2±1.0
7 February 1984	11.8±0.7	16.5±1.1	22.5±1.7	41.7±3.8
8 February 1984	12.4±0.6	17.4±0.9	25.0±1.3	42.8±2.0
8 March 1984	26.7±1.3	35.9±1.8	47.5±2.4	82.3±4.1
18 April 1984	6.2±0.3	8.8±0.4	12.5±0.7	22.4±1.1
19 April 1984	5.9±0.1	8.5±0.1	11.8±0.1	—
21 April 1984	7.0±0.1	10.4±0.3	13.8±0.2	27.2±1.8
28 November 1984	3.0±0.1	4.6±0.2	7.1±0.2	14.6±1.3
22 December 1984	4.0±0.2	5.9±0.3	9.6±0.5	17.2±0.9
22 February 1985	10.6±0.5	13.9±0.7	20.0±1.0	32.4±1.6
24 April 1985	16.5±0.2	22.4±1.1	29.8±1.5	—
19 February 1986	5.5±0.2	7.8±0.3	11.9±0.4	25.3±2.5
24 February 1986	4.7±0.2	7.8±0.3	11.6±0 4	23.5±0.8

Table 2 Near-infrared spectral indices

Date	
3 February 1983	−1.16±0.03
5 February 1983	−1.08±0.02
6 March 1983	−1.09±0.08
29 January 1984	−1.23±0.02
7 February 1984	−1.23±0.10
8 February 1984	−1.22±0.06
8 March 1984	−1.11±0.07
18 April 1984	−1.27±0.06
21 April 1984	−1.23±0.03
28 November 1984	−1.54±0.07
22 December 1984	−1.48±0.07
22 February 1985	−1.12±0.06
19 February 1986	−1.43±0.07
25 February 1986	−1.56±0.05

Because of the extreme variability of OJ287 we cannot claim to have fully sampled its temporal behaviour; however we can search for overall patterns of behaviour, which are evidence for a single origin for the emission over this wavelength range, and can lead to insights into the physical mechanism producing the variability. Even from Fig. 1 it can be seen that there is a correlation between flux and spectral index. To clarify this, Fig. 2 shows the JHKL spectral indices plotted against the log of the J-band flux. For these 14 data, performing a linear regression taking into account the uncertainty in both the flux and spectral index, we find a correlation coefficient of 0.84, representing >99% certainty of correlation. The correlation is in the sense that as the flux decreases the spectrum steepens. Conversely, as the flux increases, the spectrum flattens.

Also plotted in Fig. 2, are data from ref. 1, taken at the peak of the 1983 outburst, and refs 7, 8, taken in January to April 1983. Combining these with the present data, we find an even better correlation coefficient of 0.91. This well-defined correlation confirms the trend found for a sample of Blazars including OJ287, but with less data[6].

The power-law infrared continuum, extreme variability and high polarization[1] observed in OJ287 indicate a synchrotron origin of the infrared emission. We can explain the behaviour observed in terms of a recurrent injection or reacceleration

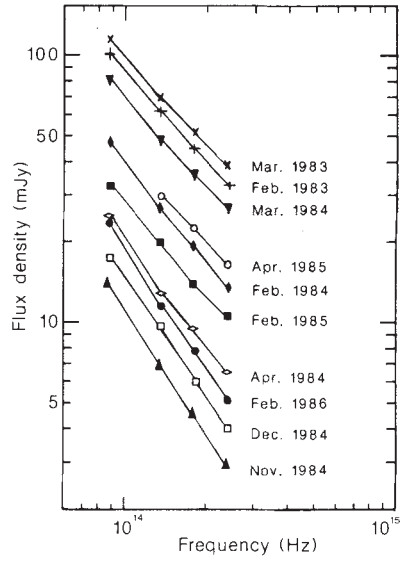


Fig. 1 The infrared spectrum of OJ287 at nine epochs over the period 1983–86. Note that none of the spectra has been displaced.

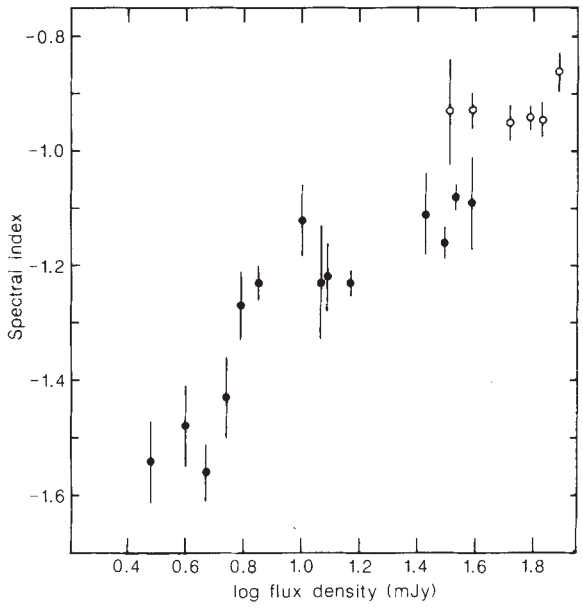


Fig. 2 A plot of infrared spectral index versus the log of the J-band flux showing a very good correlation; as the source gets fainter, the spectrum gets steeper, and vice versa. ●, JHKL spectra indices; ○, data from refs 1, 7, 8 taken at the peak of the 1983 outburst.

process. If the acceleration process produces an electron energy index of ~ -2.7 then an increase in the injection of electrons, or reacceleration of those already present, will result in an increase in emission with a spectral index ~ -0.85, such as was observed during the outburst in 1983 (ref. 1). These electrons will rapidly lose energy radiatively, however, (the results from ref. 6 suggest a magnetic field ~ 1 G in Blazars, corresponding to a lifetime measured in days for electrons radiating at 1 μm). The effective energy index will then steepen as the higher energy electrons are affected first and the observed spectrum will also steepen as it decays. In support of this conclusion we also note that there is some suggestion of a flattening in the correlation between the spectral index α_{JHKL} and $\log(S_J)$ (the log of the J-band flux) at high flux levels; this is to be expected because at high fluxes we are observing the native electron spectrum, unaffected by energy losses; the observed spectrum cannot become flatter than this.

Hanson and Coe[9] have found a similar correlation for the ultraviolet emission from OJ287 over the period 1983-84. Their ultraviolet spectral indices were generally steeper than those we see in the infrared (except when the infrared flux is very low, when the spectral index steepens to a value comparable with that seen in the ultraviolet). This implies that the ultraviolet spectrum is almost always steepened by energy losses, whereas in the infrared the true injected spectrum can sometimes be observed. We also note that from their optical UBVRI monitoring data in 1983, Moles *et al.*[4] find that the correlation between flux level and spectral index improves as one goes from the infrared into the optical region. This is probably an artefact of the spectral steepening over that wavelength range; the relative variations are larger at shorter wavelengths, due to shorter electron lifetimes, so for a limited data set the correlation will therefore show up more clearly.

The facts that the periods of increased flux after 1983 never reach the values of that outburst and that the spectral index never again reached the value of -0.85, suggest that subsequent events were not as powerful as that which caused the 1983 outburst. Continued infrared monitoring of OJ287 is essential to fully understand the emission from this extremely variable object. In particular, a long campaign of night-to-night monitoring would reveal whether this correlation is maintained on all timescales, or whether the source can take up some additional state. It would be particularly interesting to monitor OJ287 during a very low phase to discover if there was an underlying less variable component such as we have recently discovered[10] in 3C273.

OJ287 is an ideal source for this type of study, being highly variable, relatively bright and a Bl lac object so that one can directly observe the properties of the non-thermal emission over the whole spectrum. However, we note that after very careful removal of the recombination emission, Barr *et al.*[11] found a very similar correlation in the ultraviolet continuum emission from the Seyfert galaxy NGC3783. It is possible that the non-thermal emission from all active nuclei arises by a similar process, but that in many classes of object the behaviour is masked by relatively strong thermal and recombination emission.

We thank the UKIRT staff for their continued assistance with observations, particularly those obtained in service mode. W.K.G. and L.M.J.B. acknowledge financial support from SERC. UKIRT is operated by the Royal Observatory, Edinburgh, on behalf on SERC.

Received 7 July; accepted 3 October 1986.

1. Holmes, P. A. *et al. Mon. Not. R. astr. Soc.* **211**, 497-506 (1984).
2. Corso, G. *et al. Astr. Astrophys.* **140**, L23 (1984).
3. Sillinpaa, A. *et al. Astr. Astrophys.* **147**, 67-70 (1984).
4. Moles, M., Garcia-Pelayo, J. & Masegosa, J. *Mon. not. R. astr. Soc.* **211**, 621-627 (1984).
5. Gear, W. K. *et al. Astrophys. J.* **291**, 511-517 (1985).
6. Gear, W. K. *et al. Astrophys. J.* **304**, 295-204 (1986).
7. Landau, R. *et al. Astrophys. J.* **308**, 78-92 (1986).
8. Roelling, T., Becklin, E. E., Impey, C. D. & Werner, M. W. *Astrophys. J.* **304**, 646-650 (1986).
9. Hanson, C. G. & Coe, M. J. *Mon. Not. R. astr. Soc.* **217**, 831-842 (1985).
10. Robson, E. I. *et al. Nature* **323**, 134-136 (1986).
11. Barr, P., Willis, A. J. & Wilson, R. *Mon. Not. R. astr. Soc.* **202**, 453-466 (1983).

A new low-dimensional metal, $Cs[Pd(S_2C_2(CN)_2)_2]\cdot 0.5\,H_2O$

I. D. Parker*, R. H. Friend*, P. I. Clemenson† & A. E. Underhill†

* Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK
† Department of Chemistry and Institute of Molecular and Biomolecular Electronics, University College of North Wales, Bangor LL57 2UW, UK

The discovery of a low-temperature superconducting state in organic compounds of the type $(TMTSF)_2ClO_4$ ($T_c = 1.2$ K) and $(BEDT\text{-}TTF)_2AuI_2$ ($T_c = 4$ K) (where TMTSF is tetramethyltetraselenafulvalene, BEDT-TTF is bis(ethylenedithiolo)tetrathiafalvalene and T_c is the superconducting transition temperature) has stimulated the search for new materials that may show higher values of T_c (refs 1-3). The general problem encountered in molecular charge-transfer salts of this type, which have conduction bands formed by intermolecular overlap of π-electron systems, is that conduction is usually quasi-one-dimensional, with good conduction along the stacking direction. Metals with this one-dimensional character are unstable, and undergo a Peierls transition[4] to a semiconducting state at low temperatures. The relatively few exceptions (mentioned above), which remain metallic down to low temperatures, are considered to do so because they show stronger interstack interactions. We report here a new material with inherently two-dimensional interactions between the molecular π-electron systems and which we are able to stabilize as a metal down to low temperatures (1.4 K) under hydrostatic pressure (12 kbar).

Molecular metals based on square coplanar metal complexes of the KCP ($K_2[Pt(CN)_4]Br_{0.30}\cdot 3H_2O$) type have been known for the past 15 years[5]. More recently the range of compounds with metallic properties has been extended to metal-dithiolene

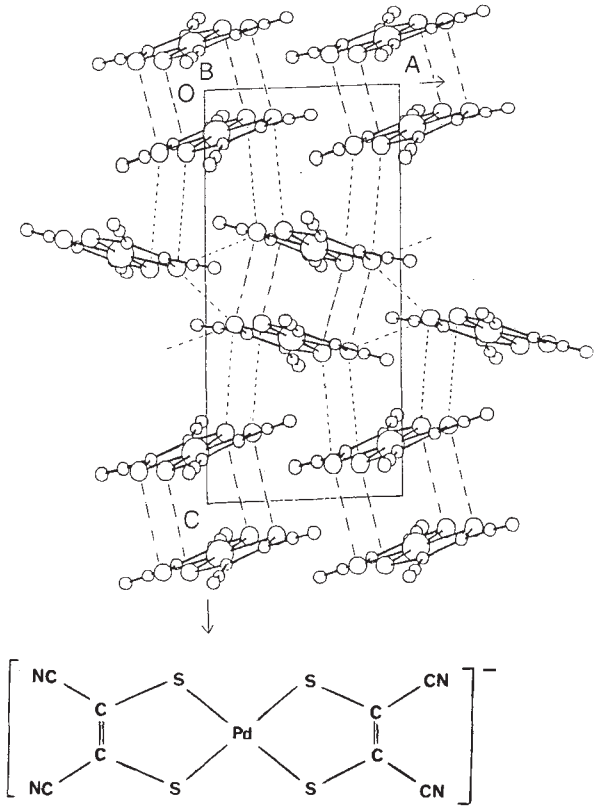


Fig. 1 *b*-Axis projection of the crystal structure of $Cs[Pd(S_2C_2(CN)_2)_2]\cdot 0.5\ H_2O$. The short S–S contacts are shown by dashed lines. Also shown is the molecular structure of $[Pd(S_2C_2(CN)_2)_2]^-$.

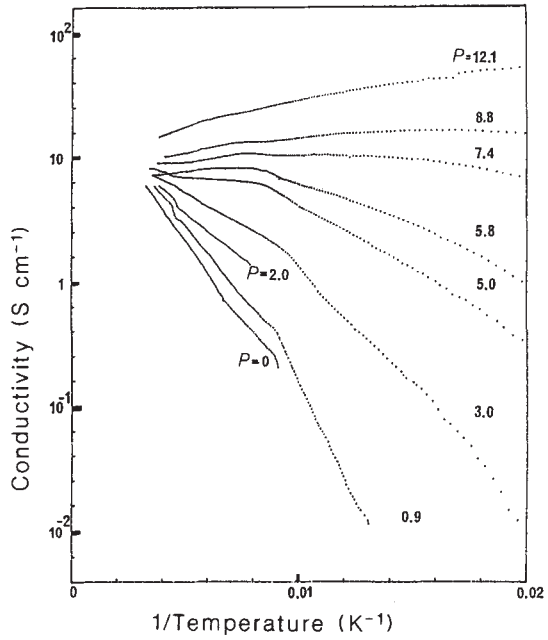


Fig. 2 Conductivity plotted against reciprocal temperature at several pressures (in kbar) for $Cs[Pd(S_2C_2(CN)_2)_2]\cdot 0.5\ H_2O$.

complexes[6]. Figure 1 shows molecular structure of the system of interest here. These are of particular interest because the intermolecular delocalization involves the π-electron systems on the ligands rather than the metal d orbitals (as in KCP). Up to now, all those molecular metals based on metal complexes have been of non-integral oxidation state; the chemical non-stoichiometry of the compound results in partial charge on the

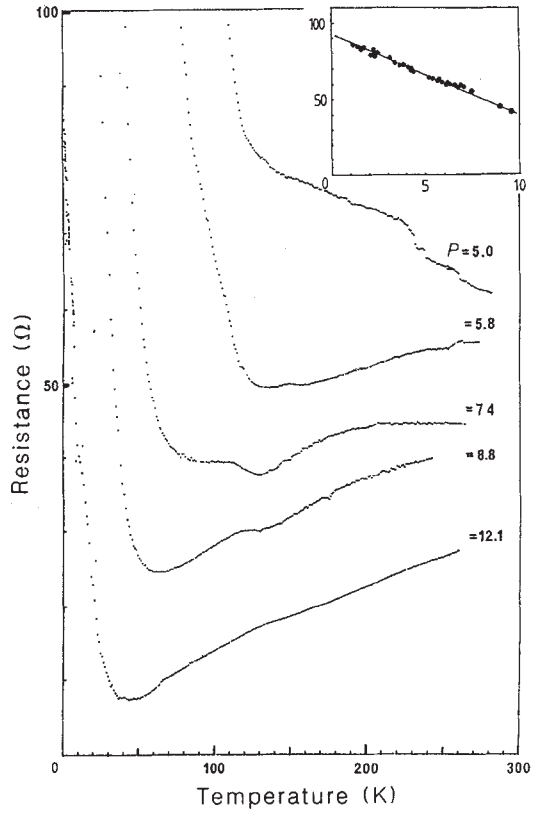


Fig. 3 Resistance plotted against temperature for $Cs[Pd(S_2C_2(CN)_2)_2]\cdot 0.5\ H_2O$. Inset shows the variation of resistance in the region 0–10 K.

complex anion, the conduction band is consequently only partly filled, and metallic behaviour results. For a stoichiometric complex with a single (integral) charge on the anion, the Peierls distortion is of the form of a simple stack dimerization. This distortion is usually very strong and metallic properties are not expected in such materials, even at room temperature. There are several examples of such semiconducting behaviour[7]. Nevertheless, in $Cs[Pd(S_2C_2(CN)_2)_2]\cdot 0.5\ H_2O$ we have discovered the first example of an integral-oxidation-state complex with metallic properties.

Black plate-like crystals of $Cs[Pd(S_2C_2(CN)_2)_2]\cdot 0.5\ H_2O$ were prepared by electrocrystallization from an aqueous solution of the dianion salt. Chemical analysis showed the compound to be stoichiometric. The crystal structure has been determined[7]; the structure consists of sheets of the $[Pd(S_2C_2(CN)_2)_2]^-$ anions separated by layers of Cs^+ and H_2O along the *b* direction. Figure 1 shows the projection of the $[Pd(S_2C_2(CN)_2)_2]^-$ anions normal to the sheet, and reveals the two-dimensional character of the structure. The $[Pd(S_2C_2(CN)_2)_2]^-$ anions form dimers in an eclipsed configuration with an intra-dimer distance of 3.329(5) Å, but these are arranged so that there are short S–S contacts between dimers, 3.807 Å along the *c* direction but only 3.576 Å along the *a* direction.

At ambient pressure this material has a room-temperature conductivity of about 5 S cm^{-1} but it is semiconducting at lower temperatures[7]. For measurements under pressure we used a Be–Cu high-pressure cell capable of operation at up to 12 kbar in the temperature range 1.4–300 K (ref. 8). Four terminal measurements of the *a* axis conductivity were made using gold paste contacts. The conductivity at room temperature rises rapidly with pressure, but the changes at lower temperatures are more dramatic, with a transition from semiconducting to metallic behaviour. Results are shown in Figs 2 and 3. Figure 2 plots conductivity against reciprocal temperature, and the

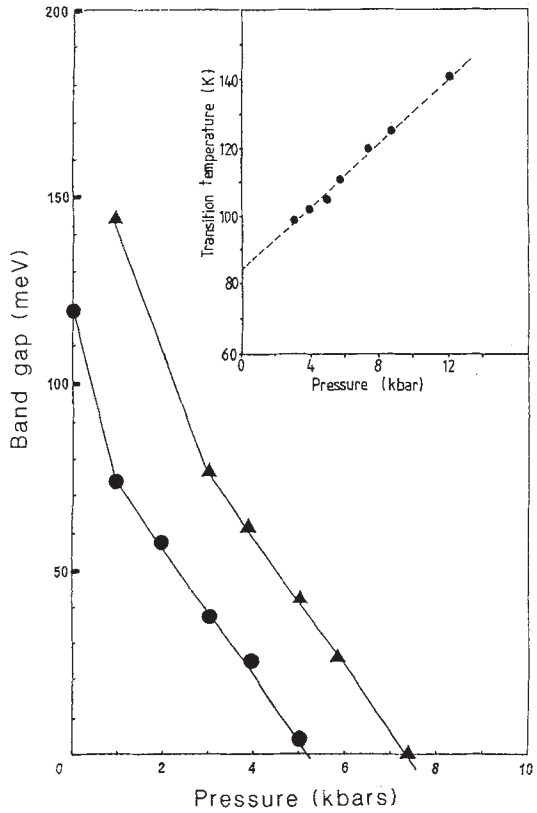


Fig. 4 Activation energies and transition temperature for the semiconducting phases of $Cs[Pd(S_2C_2(CN)_2)_2]\cdot 0.5\ H_2O$ as a function of pressure (band gap = 2 × activation energy). ●, Higher-temperature phase; ▲, lower-temperature phase.

(negative) straight-line slopes give the activation energies for conduction in the semiconducting regime. There is a monotonic decrease in activation energy to zero at about 7 kbar. Figure 3 plots resistance against temperature and shows the linear variation expected for simple metals over much of the temperature regime. At low temperatures the resistance shows an upturn; at 8.8 kbar and below this gives semiconducting behaviour. At the highest pressure investigated, 12.1 kbar, the resistance rise below 40 K is very weak, as shown more clearly in the inset to Fig. 3, and the conductivity at 1.4 K is still above 10 S cm^{-1}.

We have thus established that the conductivity can be kept at metallic levels under pressure but we need to understand why this is not so at ambient pressure. There are clues to this in the behaviour in the semiconducting regime, which is best seen in Fig. 2. There is considerable structure in the data shown, with at least two clearly defined slopes and a sharp break between them. Figure 4 shows the pressure dependence of the activation energies and of the temperature of the transition between them (chosen to be at the peak value of $\partial \ln(\text{conductivity})/\partial(1/T)$). The transitions between these phases are not of the Peierls type. For the Peierls transition there is a linear scaling between the size of the semiconducting gap and the temperature at which the transition occurs[9], whereas we see here that the transition temperature rises while the activation energies fall. We consider instead that the transitions involve some small molecular reorientations which set up appropriate superlattice potentials to 'accidentally' introduce energy gaps in the electron energy bands at the Fermi energy. We think that the effect of pressure may be to increase the intermolecular interactions to such an extent that although the structural transitions may still occur, their effect on the electron levels is diminished. Low-temperature structural measurements are required to follow this.

Received 15 August; accepted 16 October 1986.

1. Int. Conf. Synth. Metals, Kyoto, June 1986 *Synth. Metals* (in the press).
2. Williams, J. M. *Prog. inorg. Chem.* **33**, 183-220 (1985).
3. Wudl, F. *Acct. chem. Res.* **17**, 227-232 (1984).
4. Peierls, R. E. *Quantum Theory of Solids* 108-114 (Oxford University Press, London, 1955).
5. Underhill, A. E., Williams, J. M., Schultz, A. J. & Carneiro, K. in *Extended Linear Chain Compounds* Vol.1 (ed. Miller, J. S.) 73-118 (Plenum, New York, 1982).
6. Ahmad, M. M. *et al. Phys. Rev.* **B29**, 4796-4799 (1984).
7. Clemenson, P. I. *et al.* in Int. Conf. Synth. Metals, Kyoto, June 1986 *Synth. Metals* (in the press).
8. Guy, D. R. P. & Friend, R. H. *J. Phys.* E. **19**, 430-433 (1986).
9. Friend, R. H. & Jerome, D. *J. Phys.* C **12**, 1441-1477 (1980).

Organ pipe radiant modes of periodic micromachined silicon surfaces

Peter J. Hesketh*, Jay N. Zemel* & Benjamin Gebhart†

* Department of Electrical Engineering and † Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6390, USA

In recent studies on small pyroelectric thermal anemometers with roughened surfaces we showed that one of the most widely used heat transfer models[1,2] yielded calculated anemometer responses for flow and geometric behaviour that agreed functionally with observations, but were significantly smaller than the experimental data[3–5]. As the first stage in investigating the role of small structures in heat transfer, we initiated a study of emittance from deep gratings. Here we report measurements at 400 °C of infrared ($3\ \mu m \leqslant \lambda \leqslant 14\ \mu m$), normal, s- and p-polarized spectral emittances of 45 μm deep, near square-wave gratings of heavily phosphorus doped (110) silicon (P content ~ 5×10^{19} cm^{-3}). The grating surface repeat scales, Λ, were 10, 14, 18 and 22 μm, yielding a range of Λ/λ from 0.14 to 7.33. The s-polarization vector was parallel to the grating slots. Both s and p spectral emittances had pronounced resonant periodicities with a characteristic length of ~42 μm. A reasonable explanation for this behaviour is the presence of standing waves in the air slots perpendicular to the silicon surface similar to those in an organ pipe. While the resonant amplitude of the s polarization does not depend significantly on Λ it does for the p polarization. No explanation for the Λ dependence of the p polarization is known.

Although radiant emission from macroscopic black bodies and flat, smooth metallic surfaces is well understood[6,7], emission from deep, near square-wave gratings when Λ is comparable to or $<\lambda$ is not. The theory of radiant emission from structures whose scale, S, is comparable to or $<\lambda$, has drawn considerable attention[8–18]. Measurements of semiconductors emissivity gave information on the optical properties of bulk dopants[19–21]. The effect of surface asperities on metal emissivity has been thoroughly studied[22–27]. Generally, the polarized emissivity of rough metallic surfaces depends on the r.m.s. value of surface asperities with the energy concentrated at small polar angles[24–28]. The work on both s- and p-polarized photo-acoustic spectrum from a silvered bigrating is the closest to our study, but the depth (H) used was still $\ll \Lambda$ (ref. 29 and see ref. 30).

Our experiments employed (110) oriented silicon wafers on which square-wave gratings were etched to a common depth of ~45 μm using standard photolithography and a 40% (w/v) solution of KOH at 52 °C. After etching, the microconfigured wafers were spin-coated with a phosphorus based dopant and heated at 1,250 °C for 10 h. The peak phosphorus surface concentration is in the $0.4–0.6\times10^{20}$ cm^{-3} regime from sheet resistance measurements. This carrier concentration yields an effective skin depth of ≤4 μm, for $\lambda > 5$ μm. Auger measurements gave surface phosphorus concentrations of ~1×10^{20} cm^{-3}, in good agreement with the sheet resistance measurements. The heavily doped back surfaces of the silicon wafers were used as heater elements. Figure 1*a* shows the grating and the polarizations studied. Figure 1*b* shows a scanning electron micrograph of a

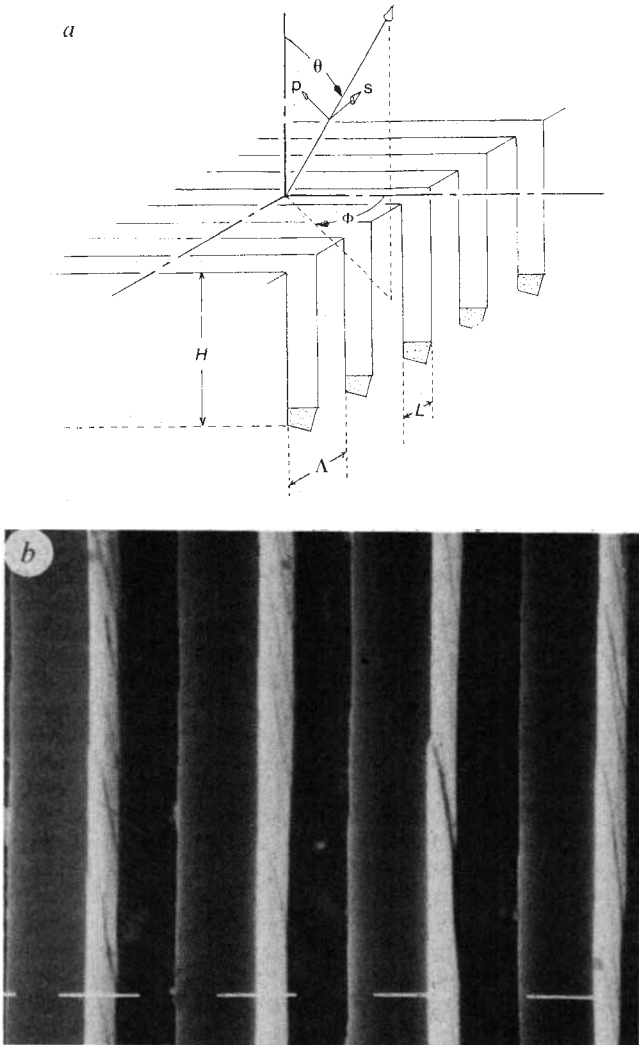


Fig. 1 *a*, Schematic drawing of the grating. Λ, repeat distance; *L*, slot width; *H*, depth; s- and p-polarization vectors; *θ*, polar angle and Φ, azimuthal angle. *b*, Electron micrograph of the microgrooves; $H = \Lambda = 22\ \mu m$, at a magnification of 1,000×. A 10-μm-wide marker is shown in white.

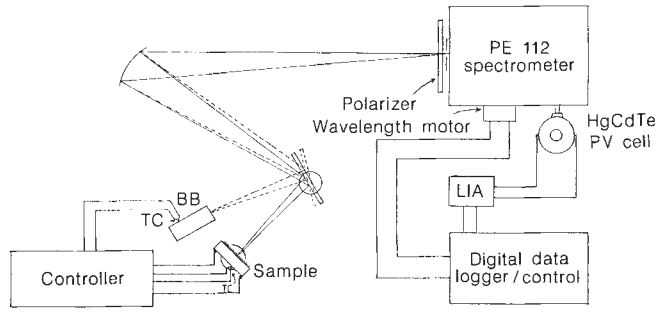


Fig. 2 Schematic layout of the emittance measuring equipment: TC, thermocouple; BB, black-body; LIA, lock-in amplifier; PE, Perkin Elmer.

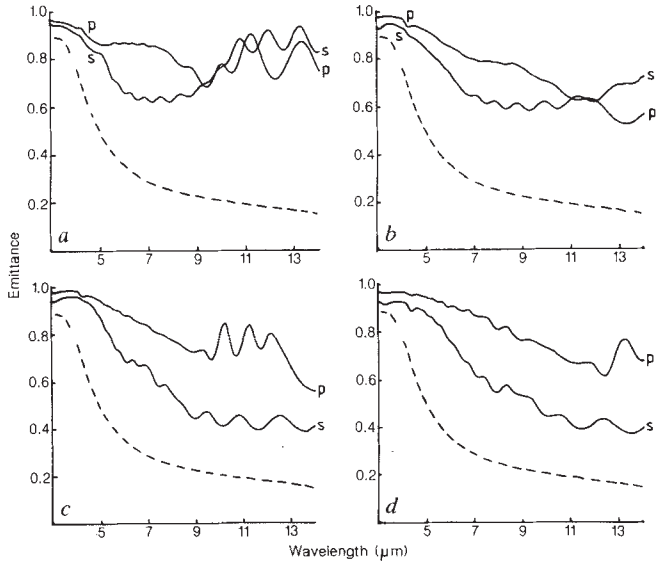


Fig. 3 Normal polarized, $T = 400\ °C$, spectral emittance as a function of spectral wavelength for the p- and s-polarizations. For comparison, the smooth surface emittance is shown as the dashed line. All gratings had a depth $H = 45 \pm 2\ \mu m$: *a*, $\Lambda = 10\ \mu m$; $L = 7.3\ \mu m$; *b*, $\Lambda = 14\ \mu m$, $L = 8.4\ \mu m$; *c*, $\Lambda = 18\ \mu m$; $L = 12.6\ \mu m$; *d*, $\Lambda = 22\ \mu m$, $L = 14\ \mu m$.

grating with $\Lambda = 22\ \mu m$. Chips (1 × 1 cm) containing the microconfigured surface were mounted in a specially designed low-thermal-loss holder. More energy was required to reach operating temperatures for microconfigured surfaced samples compared to the smooth surfaces. A black-body source was constructed from a copper cylinder. The temperature of the source and the temperature difference between it and the sample were monitored throughout the measurement sequence. The 400 °C black-body source, controlled to better than 0.01 °C, was used as a reference for stabilizing the silicon sample temperature. Temperature differences between the silicon sample and the black body were <0.5 °C. Temperature fluctuations in the sample were typically <0.1 °C during a measurement. The emissivity measurement system is shown in Fig. 2. The signal-to-noise ratios of the infrared detector were ~100:1 or better at the higher temperatures. Data obtained at lower temperatures agreed well with these observations. The data were processed with a small data logging system based on a single board Z-80 computer.

The measurement procedure consisted of logging the polarized spectral intensity of the 400 °C black-body and silicon grating, both at normal incidence. This corresponded to a polar angle, $\theta = 0°$, and an azimuthal angle, $\Phi = 90°$. The polarized spectral emittance, $\varepsilon(^{s}_{p}; \lambda; T = 400\ °C)$ is defined as the ratio of these spectral intensities. For the above angles, the s-polarization vector is parallel and the p-polarization is perpendicular to the slots (Fig. 1*a*). Shifting Φ to 0° reverses the definitions of the s- and p-polarization vectors relative to the slots so they are perpendicular and parallel to the slots, respectively. The reproducibility of ε was 2-4% between runs. Pronounced periodic oscillations were found in both $\varepsilon(p; \lambda; 400\ °C)$ and $\varepsilon(s; \lambda; 400\ °C)$ from gratings with $\Lambda = 10\ \mu m$, 14 μm, 18 μm, and 22 μm (Fig. 3). The value of ε for a smooth, heavily doped, silicon surface is shown in Fig. 3 as a dashed curve. Temperature had little influence on the magnitude of ε and essentially none on the period or amplitude of the observed oscillations.

The wavenumber, $K^{(i)}$, corresponding to maxima in the polarized spectral emittance ($K^{(i)}_m = 1/\lambda^{(i)}_m$, $i = s, p$) were plotted against an integral mode number, $m^{(i)}$ (for the appropriate polarization, $i = s, p$). The origin of the $K^{(i)}_m$ plot for the individual gratings was adjusted so that the data overlapped as shown in Fig. 4. Note that there is essentially no influence of Λ on the slope. We constructed a model in which the $K^{(i)}_m$ are related to the $m^{(i)}$ by

$$K^{(i)}_m = \begin{cases} m^{(i)}/2H \\ (2m^{(i)}+1)/4H \end{cases} \quad (1)$$

where the upper relation applies to modes where the nodes

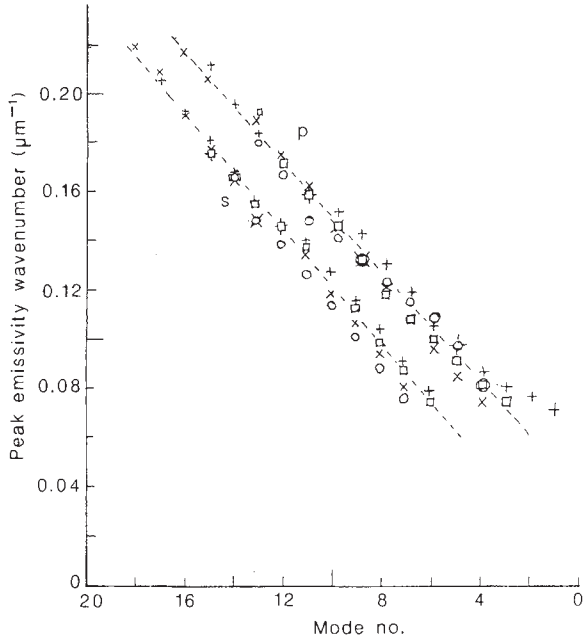


Fig. 4 Peak emittance wavenumber versus mode number; p- and s-polarizations; $H = 45 \pm 2$ μm; Λ, ○ = 10 μm; ⊙ = 14 μm; + = 18 μm; × = 22 μm.

occur at 0 and H and the second where one node occurs at 0 and the anti-node is at H. In both cases,

$$\Delta K_m^{(i)} = \Delta m^{(i)}/2H \quad (2)$$

From Fig. 4, $H = 42$ μm in striking agreement with the average depth of ~45 μm measured for the four different gratings. This is even more remarkable because Λ varies by a factor of >2. Whereas the slope of $K_m^{(i)}(m^{(i)})$ is independent of Λ, inspection of the amplitudes of the emittance oscillation of Fig. 3 suggests that the p-polarized data oscillations are far more sensitive to Λ than are the s-polarized. We do not yet have any explanation for this difference.

The s-polarized emittance oscillations have no parallel in either experiment or theory. Work by Inagaki *et al.* suggests that periodic resonant absorptance occurs on a bigrating, but their data indicate that the magnitude of the absorptance is much smaller for the s-polarization than for the p-polarization[29]. Recent calculations show that substantial field enhancement arises in the slots of a deep cavity for the s-polarizations[14,18]. However, the far-field pattern associated with these s-polarized radiative modes is weak and almost identical to those from a smooth mirror of similar composition. In our structures, thermal energy populating the radiative (or absorptive) modes of the deep grating structure may be coupled out by asperities in the slot geometries. In any case, the peaks in the emittance are due to emission from 'organ pipe' or planckian type resonant modes in the slots of the grating.

We thank Judson Infrared for the HgCdTe infrared detector, Dr A. Ipri for supplying the photomasks; the EEMS Program of the NSF for partial support, D. Silversmith, Program Manager, under grant ECS-84-12241 (P.H., J.N.Z.); partial support from NSF grant CBT-84-18517 (B.G.); and Professors H. Baltes and A. Maradudin for useful discussions.

Received 10 June; accepted 10 September 1986.

1. Schlichting, H. *Boundary Layer Theory* (McGraw Hill, New York, 1968).
2. Gebhart, B. *Heat Transfer* (McGraw Hill, New York, 1961).
3. Rahnamai, H. & Zemel, J. N. *Sensor Actuat.* **2**, 1-16 (1981).
4. Frederick, J. R., Zemel, J. N. & Goldfine, N. F. *J. appl. Phys.* **57**, 4936-4943 (1985).
5. Hesketh, P., Zemel, J. N. & Gebhart, B. *J. appl. Phys.* **57**, 4944-4949 (1985).
6. Planck, M. *Theory of Heat Radiation* (Eng. trans. of *Theorie der Warmestrahlung*, 2nd edn 1913) (Dover, New York, 1959).
7. Siegel, R. & Howell, J. R. *Thermal Radiation Heat Transfer* 2nd edn (McGraw Hill, New York, 1981).
8. Baltes, H. P. *Infrared Phys.* **16**, 1-8 (1976).
9. Derrick, G. H., McPhedran, R. C., Maystre, D. & Neviere, M. *Appl. Phys.* **18**, 39-52 (1979).
10. Elson, J. M. & Sung, C. C. *Appl. Opt.* **21**, 1496-1501 (1982).
11. Chang, R. K. & Furak, T. E. (eds) *Surface Enhanced Raman Scattering* (Plenum, New York, 1982).
12. Fano, U. *J. opt. Soc. Am.* **31**, 213-222 (1941).
13. Hessel, A. & Oliner, A. A. *Appl. Opt.* **4**, 1275-1297 (1965).
14. Wirgin, A. & Maradudin, A. A. *Phys. Rev.* **B31**, 5573-5576 (1985).
15. Maystre, D. in *Electromagnetic Surface Modes* (ed. Boardman, A. D.) 166 (Wiley, New York, 1982).
16. Glass, N. E., Maradudin, A. A. & Celli, V. *Phys. Rev.* **B27**, 5150-5153 (1983).
17. Garcia, N. & Maradudin, A. A. *Opt. Commun.* **45**, 306-310 (1983).
18. Wirgin, A. & Maradudin, A. A. *Appl. Surf. Sci.* (submitted).
19. Stierwalt, D. L. & Potter, R. F. *Proc. int. Conf. on the Physics of Semiconductors, Exeter*, 513 (Institute of Physics and the Physical Society, London, 1962).
20. Stierwalt, D. L. *Appl. Opt.* **5**, 1911-1915 (1966).
21. Stierwalt, D. L. & Potter, R. F. in *Semiconductors and Semimetals* Vol. 3 (eds Willardson, R. K. & Beer, A. C.) (Academic, New York, 1968).
22. Lofving, S. *Appl. phys. Lett.* **36**, 632-633 (1980).
23. Edwards, D. K. *Advances in Thermophysical Properties at Extreme Temperatures and Pressure, 3 int. Symp.* 189 (ASME, New York, 1965).
24. Kanayama, K. *Heat Transfer—Jap. Res.* **1**, 11-21 (1972).
25. Rolling, R. E. *Progress in Astronautics and Aeronautics: Thermophysical Properties of Spacecraft and Planetary Bodies* (ed. Heller, G. B.) (Academic, New York, 1967).
26. Craig, S. & Harding, G. L. *Sol. Energy Mater.* **4**, 245-264 (1981).
27. Petit, G. D., Cuomo, J. J., DiStefano, T. H. & Woodall, J. M. *IBM J. Devl.* **22**, 372-377 (1978).
28. Church, E. L., Jenkinson, H. A. & Zawada, J. M. *Opt. Engng.* **16**, 380-385 (1977).
29. Inagaki, T. Goudennet, J. P., Little, J. W. & Arakawa, E. T. *J. opt. Soc. Am.* **B2**, 433-439 (1985).
30. Baltes, H. P. & Hilf, E. R. *Spectra of Finite Systems* (Bibliographisces Institute, Zurich, 1976).

Evidence for a link between atmospheric thermonuclear detonations and nitric acid

G. Holdsworth

National Hydrology Research Institute and
Arctic Institute of North America, University of Calgary,
2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4

Suitably located glacier cores, obtained from high-altitude, low-temperature sites, can reveal detailed information about atmospheric air chemistry at sub-annual resolution[1]. Such data may provide input to climate-change models, the study of acid precipitation patterns and many other phenomena. Here I present data from an ice core which show that during the era of intense atmospheric thermonuclear weapons testing (ATWT) a significant part of the nitrate content in the snow was modulated by the intensity of the nuclear detonations. The fixation of nitrogen by nuclear fireballs leads to NO_x gases in the atmosphere[2] and ultimately to nitric acid in precipitation. At certain concentrations, these gases and the associated aerosols may perturb the climate[3,4].

In 1980 a series of snow and ice cores were retrieved from an altitude of 5340 m (approximately the 500 mbar or mid-troposphere level) on the upper plateau of Mt Logan (60° 36′ N; 140° 30′ W) in the Yukon Territory, Canada. The recorded mean annual snow temperature is −28.5 °C and the mean net annual accumulation rate is ~0.37 m water equivalent. The gross β-activity levels in the upper part of one of these cores were determined to a depth of almost 27 m, corresponding to about the year 1943 (ref. 1). These data were interpreted in terms of the known chronology of ATWTs[1,5].

Analyses for the major ions in the firn were performed at the Laboratoire de Glaciologie, CNRS, Grenoble, for a section of core between 10 and 21 m (AD 1967-1953)[6]. Inspection of the nitrate concentration data and the gross β-activity curve showed that these two data sets shared several major features: a prominent 1963 maximum preceded by a ~1960 (temporary moratorium) minimum and followed by another (post-1962 limited test ban treaty) minimum in ~1966. Analyses of the upper 10 m of core, with overlap on the lower section, were then carried out at the University of Calgary, using the same precautions exercised earlier[6,7]. For these purposes a parallel

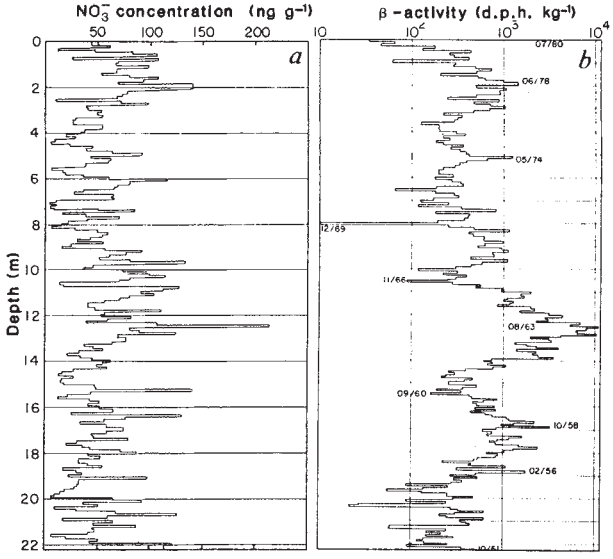


Fig. 1 *a*, Raw data for nitrate concentration and *b*, gross β-activity levels plotted against depth in the core. The top of the core corresponds to mid-1980 and the bottom to ~1952. Prominent peaks and troughs in *b* have dates assigned according to ref. 1.

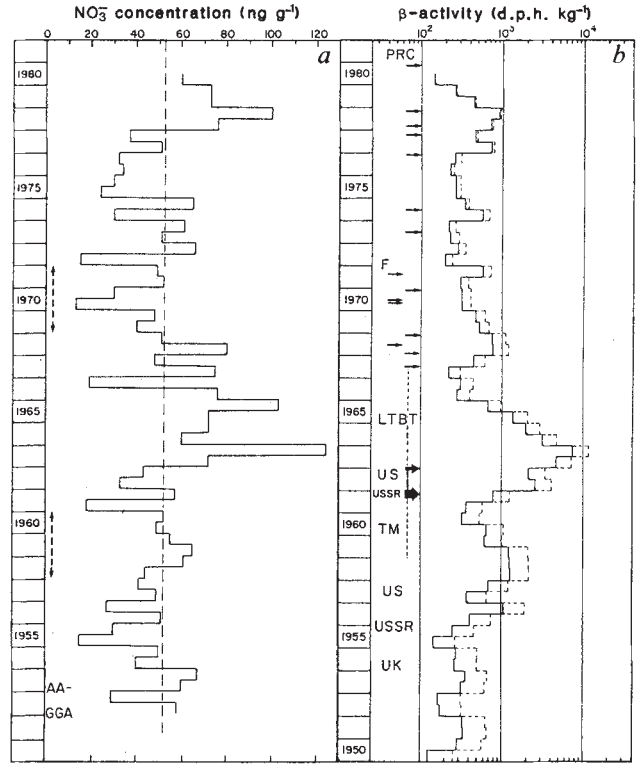


Fig. 2 Time-series plots of *a*, nitrate concentration, where broken arrows indicate position of auroral activity (AA) or global geomagnetic activity (GGA) maxima[20] and *b*, gross β-activity levels, where the arrows indicate the dates of prominent atmospheric thermonuclear weapons testing (PRC, People's Republic of China; F, France; US, United States; USSR, Union of the Soviet Socialist Republic; UK, United Kingdom; TM, Temporary Moratorium; LTBT, Limited Test Ban Treaty). The dashed blocks above the solid blocks indicates an approximate reconstruction of the original gross β-activity levels at the time of deposition.

core taken within 1 m of the earlier core[6] was analysed, thus providing a check on the consistency of results. At the 0.10 m cutting interval used, spikes in the nitrate and other anion data are quite common and cannot be associated with contamination. Contamination is further discussed in the section dealing with the origin of nitrates in the core. A comparison was then made between the upper 10 m of nitrate data and the gross β-activity curve (Fig. 1). This comparison revealed that, at least from 1973 to 1980, the People's Republic of China ATWTs were tending to modulate the NO_3^- levels.

Because the Chinese and the preceding Soviet ATWTs were done at locations with the same general latitude (±15°) as Mt Logan, and because the latter represents a relatively near-field site, it is not unexpected that the chemistry of the core should be influenced by these ATWT events. It has already been shown[8] that ATWT events produce copious quantities of nitric oxide (NO) ($\sim 10^{10}$ g Mt^{-1}). Although most of the detonations were made in the troposphere, large quantities of NO were injected, with other explosion products, into the stratosphere, where the residence time of the gases and aerosols could be up to several years[8].

A timescale for the core was established[1] to an accuracy of between ±0.25 yr and ±0.5 yr. The original data (typically 10 samples per year) were then converted into 6-monthly mean values to produce the time series of gross β-activity and $[NO_3^-]$ shown in Fig. 2. This particular time interval used for averaging represents the shortest unit that can be justifiably used while still preserving significant structure in the time series. Further smoothing was carried out for cross-correlation purposes. An attempt was made to reconstruct the gross β-activity values to their original values at the time the snow was deposited. This would allow a closer comparison to be made between gross β-activity and $[NO_3^-]$ with respect to amplitude information. The dashed blocks in Fig. 2*b* correspond to a simple radioactive decay correction based on the assumption that the dominant radionuclides present were ^{90}Sr and ^{137}Cs. This assumption is evidently not a good approximation for the most recent 8 yr of data[9], but reconstructions made using a sample-age-dependent effective half life[9] yield results at variance with the gross β-activity data[1] for air. Errors introduced by this adjustment of the data will not materially affect the major conclusions of this paper.

Cross-spectral analyses were performed on the two time series. From the measured values of the gross β-activity, A, the value of the cross-correlation coefficient, r, between the two series is +0.42 at zero lag and +0.45 to +0.48 at 2–2.5 yr lag. From the decay-corrected gross β-activity values, A_0 the corresponding values of r are +0.41 and +0.44 to +0.47 respectively, all significant at the 95% level. For other lags the correlation is not significant. The coherence at zero lag was expected, as the two time series exhibit this feature visually. A clear explanation for the apparent correlation at 2–2.5 yr lag cannot be offered. It may represent a statistical artefact.

Several studies[10–13] concerned with the origin of nitrate in snow, firn and ice cores give twelve possible sources of this major trace ion, including atmospheric thermonuclear explosions. Local contamination of the core is not considered applicable here. Strict procedures and controls employed during sampling and analysis ensured that nitrate contamination was not significant. Measurements were made at two independent research institutions and a third commercial laboratory as a check by two different methods (ion chromatography and colorimetry). The core was independently sampled for gross β-activity, and as that curve resembles quite closely other published curves from different regions of the world[1,5], it is concluded that contamination cannot be significant. Finally, variations in nitrate in the core may also result from changes in accumulation rate[12], ablation, or meteorological parameters not associated with these quantities.

It is not possible to quantify each of the above sources in a detailed manner and it is furthermore evident that relative contri-

butions from these sources vary geographically, both regionally[7,10-12] and with altitude[13,14]. It was shown[7] that no detectable trends in either $[SO_4^{2-}]$ or $[NO_3^-]$ exist in the Mt Logan snow from ~1880 to 1980, thus evidently eliminating anthropogenic input related to fuel combustion, as interpreted from ice cores from lower-altitude Greenland and the NE Canadian Arctic[15-17]. Therefore the most likely sources for significant nitrate input to Mt Logan snow in the specific time interval considered here are: (1) ionization in the upper atmosphere from various sources[10,11], (2) meteoroid fixation[10,11], (3) photochemical fixation[10,11,18], (4) volcanic emissions[7,19] and (5) atmospheric thermonuclear weapons testing[8]. The one mechanism for producing ionization in the upper atmosphere for which there is an established cyclicity (sunspot-auroral activity) does not appear to produce a signal in these data or in the Greenland data[12]. In Fig. 2*a* the broken arrows indicate periods of maximum auroral and geomagnetic activity[20]. No data are available for evaluating sources (2) and (3). In addition, the period AD 1952-80 was not particularly marked by major volcanic eruptions close enough to Mt Logan to cause a significant $[NO_3^-]$ pulse, unlike in 1912 for Novarupta[7]. Most of the relatively large eruptions that occurred in the north Pacific during this period took place on the Kamchatka Peninsula[19]. There is no consistent correlation between these events and the peaks in $[NO_3^-]$ except for Kliuchevskoi (1965, 66) which has a volcanic explosive index (VEI) of 3, whereas others during this period had VEI values of 4 and 5 (ref. 19). However, gaseous emissions are not necessarily related to VEI[21,22]. There were other volcanic events[19] that occurred during the 1964-66 interval that could have contributed cumulatively to the significant peak in $[NO_3^-]$ for AD 1965 ± 0.5, which is particularly anomalous with respect to the ATWT history. Of course, any input of nitrate from these volcanic sources depends quite crucially on air-mass trajectories. A sulphate peak occurs with the nitrate peak for 1965 (ref. 6) so at least some volcanic input is likely.

The generation of NO from ATWT explosions has been put on a semi-quantitative basis[8], thus providing an opportunity for checking whether the core data are consistent with theoretical estimates. The 1963-65 'excess' nitrate is defined as that amount above the mean for 1952-80, or 52 ng g^{-1} (Fig. 2*a*). It will be assumed initially that the 1965-66 spike, which does not correlate with an enhanced gross β-activity level, is due to ATWTs: in 1965 and 1966 the People's Republic of China conducted small ATWTs. This assumption would imply different transport processes or routes for the two atmospheric species being considered. Next, the excess nitrate is converted into a load value by using the net annual snow accumulation data for these years. This calculation yields a total value of 4.45×10^4 $\mu g\ m^{-2}$ of NO_3^-, or roughly 4.5×10^4 $\mu g\ m^{-2}$ of HNO_3. If this load density is assumed to apply uniformly over the Northern Hemisphere (2.54×10^{14} m^2) then a total of about 11 Tg of HNO_3 is obtained. This is surely an upper limit, because if most of this HNO_3 is derived from ATWT events originating in the USSR, then the load density at the Mt Logan site must be much higher than the average for the Northern Hemisphere, as it is for the gross β-activity[1]. Because it was found that the gross β-activity at Mt Logan was about 32% higher than in Greenland[1], it is reasonably assumed that the Mt Logan nitrate flux from ATWTs is also correspondingly higher than in Greenland. If the Mt Logan nitrate flux derived from ATWTs is assumed to be twice the northern hemispheric average, then the total load for the stated period would be reduced to about 6 Tg. If, now, the excess nitrate during 1965 and up to 1966 is due to local volcanic emissions, then by trimming off the 1965-66 spike to the long-term mean level, a revised total computed load would be close to 5 Tg.

Assuming that a 1 megatonne bomb yield produces $\sim 10^{10}$ g NO, equivalent to $\sim 2 \times 10^{10}$ g HNO_3, and that 340 megatonnes of cumulative yield occurred between October 1961 and December 1962 (ref. 8), then the total yield of HNO_3 for this period would be ~6 Tg. Another estimate[2] of cumulative bomb yield for this period is <340 Mt, so that a lower nitrate load is possible. From these rough calculations it may be seen that the excess nitrate for 1963-65, as derived from Fig. 2*a*, is of the right order for it to be accounted for by the known ATWT events that occurred during the preceding period.

In conclusion, the data presented here suggest that the ATWT events of more than two decades ago generated significant quantities of nitric oxide, the precursor gas for the nitric acid found in the snow on Mt Logan. The all-time peak pulse in gross β-activity coincides with the largest nitrate pulse for the period 1952-80. The observed excess nitrate, above the three decades mean for the 1963-65 period, can roughly be accounted for if the theoretical predictions for ATWT-produced NO are correct. These results, together with the observation that the gross β-activity curve is similar to other curves covering the same period, from different parts of both hemispheres[5], strongly suggest a common physical connection between nitrate and the radionuclides responsible for the β-activity in the snow, rather than just a connection through a common transport process. This latter mechanism would almost certainly explain the significant coherence between $[SO_4^{2-}]$ and $[NO_3^-]$ found particularly in the lower 10 m section of the core studied. For these two species $r = +0.49$ at zero lag, significant at the 95% level. Because this applies during the interval that includes the heaviest ATWT detonations, it would appear that the radioactive particles were playing a dominant role in acting as the nuclei for the aerosols that carried both the H_2SO_4 and HNO_3, in accordance with stratospheric aerosol theory[4]. A possible mechanism for injecting the stratospheric aerosols into the troposphere could be tropopause folding[23], a process that may be enhanced by lower-stratosphere overloading. Lastly, more direct evidence suggesting a connection between the 1976 People's Republic of China ATWT to a pulse of atmospheric nitrate in filtered air samples has been reported (W. A. Sedlacek, E. J. Mroz, A. L. Lazrus and B. W. Gandrud, in preparation).

I thank P. Upton and A. Williams, and others associated with the Arctic Institute of North America; the ice core drilling team who made the recovery of the core possible; R. Delmas and W. Sedlacek for helpful discussions; B. Gosling and B. Smith for the statistical processing of the data; and R. VanEverdingen and J. Winchester for significant improvements to the manuscript, a preliminary version of which was presented at the 1985 AGU fall meeting in San Francisco. This paper is published with the approval of the Director, National Hydrology Research Institute, Canada.

Received 7 April; accepted 8 October 1986.

1. Holdsworth, G., Pourchet, M., Prantl, F. A. & Meyerhof, D. P. *Atmos. Envir.* **18**, 461-466 (1984).
2. Carter, M. W. & Moghissi, A. A. *Hlth Phys.* **33**, 55-71 (1977).
3. Turco, R. P., Whitten, R. C. & Toon, O. B. *Rev. Geophys. Space Phys.* **20**, 233-279 (1982).
4. Prospero, J. M. *et al. Rev. Geophys. Space Phys.* **21**, 1607-1629 (1983).
5. Pourchet, M., Pinglot, F. & Lorius, C. *J. geophys. Res.* **88**, 6013-6020 (1983).
6. Delmas, R., Legrand, M. & Holdsworth, G. *Ann. Glaciol.* **7**, 213 (1985).
7. Holdsworth, G. & Peake, E. *Ann. Glaciol.* **7**, 153-160 (1985).
8. Foley, H. M. & Ruderman, M. A. *J. geophys. Res.* **7**, 4441-4450 (1973).
9. Ambach, W. & Rehwald, W. *Arct. alpine Res.* **14**, 163-166 (1982).
10. Parker, B. C., Zeller, E. J. & Gow, A. J. *Ann. Glaciol.* **3**, 243-246 (1982).
11. Laird, C. M., Zeller, E. J., Armstrong, T. P. & Parker, B. C. *Geophys. Res. Lett.* **9**, 1195-1198 (1982).
12. Herron, M. M. *J. geophys. Res.* **87**, 3052-3060 (1982).
13. Lyons, W. B. & Mayewski, P. A. *Geophys. Res. Lett.* **10**, 1160-1163 (1983).
14. Bollinger, M. J. *et al. J. geophys. Res.* **89**, 9623-9631 (1984).
15. Koerner, R. M. & Fisher, D. A. *Nature* **295**, 137-140 (1982).
16. Neftel, A., Beer, J., Oeschger, H., Zurcher, F. & Finkel, R. C. *Nature* **314**, 611-613 (1985).
17. Mayewski, P. A. *et al. Science* **232**, 975-977 (1986).
18. Legrand, M. R. & Delmas, R. J. *Atmos. Envir.* **18**, 1867-1874 (1984).
19. Simkin, T. *et al. Volcanoes of the World* (Hutchinson Ross, Stroudsburg, Pennsylvania, 1981).
20. Legrand, J. P. & Simon, P. A. *Solar Phys.* **70**, 173-195 (1981).
21. Sedlacek, W. A., Mroz, E. J., Lazrus, A. L. & Gandrud, B. W. *J. geophys. Res.* **88**, 3741-3776 (1983).
22. Berresheim, H. & Jaeschke, W. *J. geophys. Res.* 3732-3740 (1983).
23. Shapiro, M. A. *J. atmos. Sci.* **37**, 994-1004 (1980).

Weakening of rock salt by water during long-term creep

Janos L. Urai, Christopher J. Spiers, Hendrik J. Zwart & Gordon S. Lister*

Institute of Earth Sciences, University of Utrecht, Budapestlaan 4, 3508 TA Utrecht, The Netherlands

The rheological properties of rock salt are of fundamental importance in predicting the long-term evolution of salt-based radioactive waste repositories and strategic storage caverns, and in modelling the formation of salt diapirs and associated oil traps[1,2]. The short-term, high-stress rheology of rock salt is well known from laboratory experiments; however, extrapolation to appropriately low stresses fails to predict the rapid flow seen in certain natural structures. Furthermore, experiments have failed to reproduce the recrystallized microstructure of naturally deformed salt. Here we report experiments indicating that the above discrepancies can be explained by taking into account the influence of trace amounts of brine. Trace brine is always present in natural salt but sometimes escapes during experiments. Our tests on dry dilated salt show more or less conventional dislocation creep behaviour, but brine-bearing samples show marked weakening at low strain rates. This is associated with dynamic recrystallization and a change of deformation mechanism to solution transfer creep. Because natural rock salt always contains some brine, these results cast substantial doubt on the validity of presently accepted dislocation creep laws for predicting the long-term rheological behaviour of salt in nature.

Creep of rock salt has been the subject of extensive investigations over the past two decades[1]. Existing data[3,4] show good agreement for a wide range of rock salts, and satisfactorily describe short-term mine convergence[5]. Long-term rheological properties, however, are not accessible in the laboratory, and predictions are generally based on extrapolation. All previous experimental work indicates that 'dislocation creep' is the most important deformation mechanism in salt. This mechanism also seems common during natural deformation, as evidenced by preferred lattice orientations[1,6] and the presence of dislocation substructures[7]. However, many naturally deformed salts also show evidence of (dynamic) recrystallization by grain-boundary migration[8,9], a process rarely observed in experiments[6,10] and generally held to be impossible under natural conditions[1,11,12]. This discrepancy between the microstructure of naturally and experimentally deformed salt casts doubt on the applicability of laboratory flow data to natural deformation, as it suggests a change in microscale processes at low strain rates.

The flow rates of salt glaciers increase dramatically after a few centimetres of rainfall[13,14], to rates several orders of magnitude faster than suggested by experimental data, and the salt apparently undergoes extensive recrystallization. On this basis, it has been proposed that small amounts of brine can dramatically alter the deformation mechanisms and rheology of rock salt[13,15]. As all natural rock salts contain small amounts of brine (usually 0.1-1.0 wt.%)[16], the weakening effects of water operating in glacier salt should also be important in other salt deposits[13,14].

On theoretical grounds[8], even very small amounts of brine can be expected to influence salt creep. First, dynamic recrystallization can be strongly enhanced by thin brine films at grain boundaries[8,17]; secondly, such films can cause solution-transfer creep[8,17-20]; thirdly, water can diffuse into the sodium chloride lattice[21,22], possibly affecting dislocation dynamics. To investigate these possibilities, we conducted two series of experiments.

In the first series, we used natural rock salt samples from the Asse Mine (West Germany). The material consists of a very pure (99%) halite rock, with a grain size of ~5 mm (Fig. 3*a*) and a dynamically recrystallized microsctructure[8,9]. The water content is ~0.05 wt%[23], up to half of which is present in fluid inclusions at grain boundaries (Fig. 3*b*, *i*). Right-cylindrical samples, 70 mm in diameter and 150 mm long, were machined from this material, and deformed at 150 °C at a constant strain rate of $\sim 10^{-5}\ s^{-1}$, followed by stress relaxation[8,24]. A dilatometric triaxial testing machine was used[8]. The pre-relaxation stress-strain behaviour was similar for all samples (Fig. 1), but clear differences were observed in dilatancy and stress relaxation (Figs 1, 2). Two types of behaviour can be distinguished: (1) Samples deformed at a confining pressure of

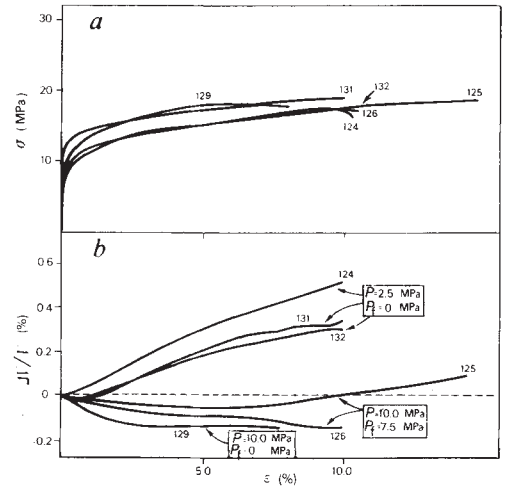


Fig. 1 *a*, Differential stress (σ) and *b*, volume change ($\Delta V/V$) plotted against strain (ε) for the pre-relaxation stage of the deformation experiments on natural (Asse) samples. $T = 150$ °C; $\dot{\varepsilon} = 3\times10^{-5}\ s^{-1}$. The confining pressure (P) and pore fluid (brine) pressure (P_f) used in each test are indicated. Note that samples deformed at $P = 2.5$ MPa show substantial dilatancy (volume increase) compared with those deformed at $P = 10$ MPa.

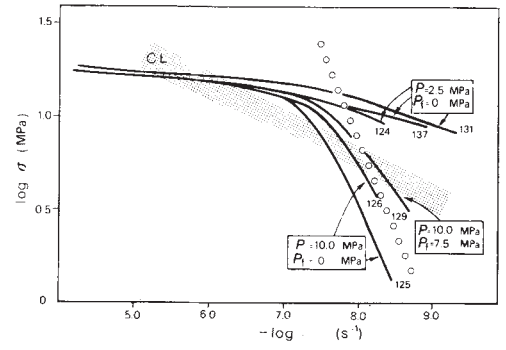


Fig. 2 Differential stress (σ) plotted against strain rate ($\dot{\varepsilon}$) (solid curves) for the stress relaxation stage of the experiments shown in Fig. 1. Note the dramatic weakening effect seen in samples deformed in non-dilatant conditions ($P = 10.0$ MPa), at strain rates of $<10^{-7}\ s^{-1}$. This low-strain-rate behaviour can be described by a power-law creep equation, $\dot{\varepsilon} = A\sigma^n$, where the stress exponent (or stress sensitivity of strain rate, n) has a value of 1-2. Samples deformed in dilatant conditions ($P = 2.5$ MPa) are stronger than predicted by previously existing creep laws[1,3,4,25] (stippled area, Cl) but exhibit similar stress sensitivity (n values) at the lower strain rates. Note that the weakening behaviour observed in non-dilatant conditions ($P = 10.0$ MPa) agrees well with predictions based on the solution-precipitation creep model given in the text (equation (1)). This model was applied to the recrystallized samples using $d = 3$ mm, $a = 44$ (ref. 18) and $C = 8\times10^{-8}$ m (ref. 8). The diffusivity parameter B was estimated from published data[8].

* Lamont-Doherty Geological Observatory, Palisaes, New York 10964, USA.

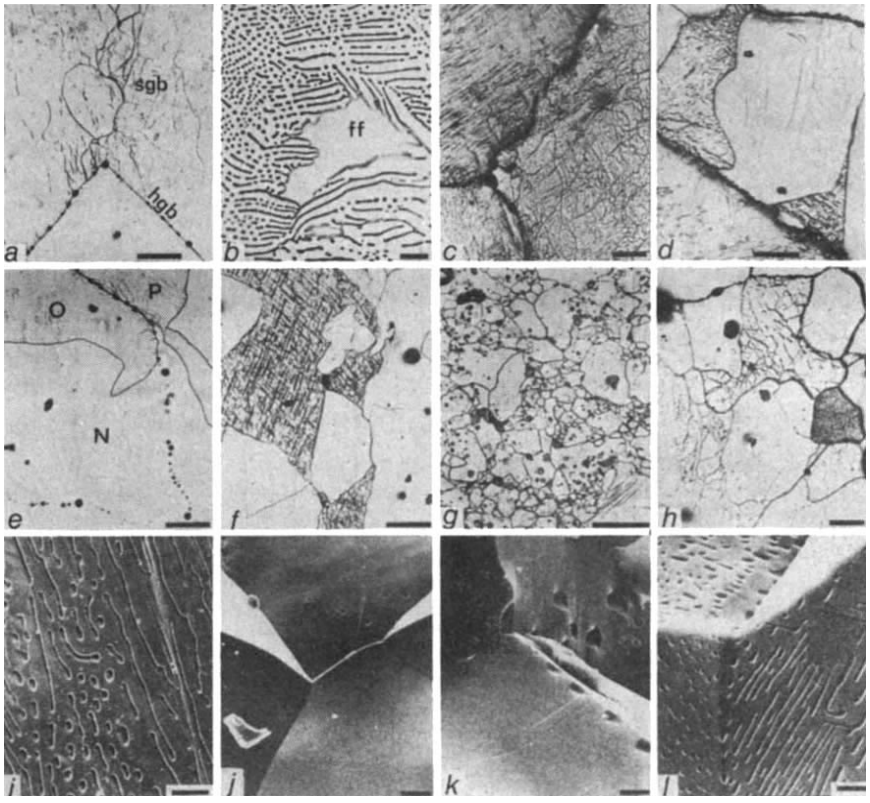


Fig. 3 *a*, Reflected-light micrograph of a polished and etched section of the natural starting material (from the Asse Mine) used in the present experiments. A grain containing subgrain boundaries (sgb) with a few degrees misorientation across them can be seen to be replaced by a newer grain, with fluid inclusions developed in the high-angle grain boundary (hgb) separating the two grains. Scale bar, 0.1 mm. *b*, Transmitted-light micrograph of the Asse salt shown in *a*, showing the typical structure of a high-angle boundary in plan view (see *i*). Note the array of fluid inclusions and fluid film segments (ff). These generally contain gas and/or brine. Scale bar, 0.1 mm. *c*, Reflection micrograph showing typical microsctructure of Asse salt samples deformed in dilatant conditions (P = 2.5 MPa). Note the densely developed slip-band structures, crude subgrains and dilated grain boundaries. Scale bar, 0.1 mm. *d*, Reflection micrograph illustrating the typical recrystallized microstructure of Asse samples deformed in non-dilatant conditions (P = 10.0 MPa). The micrograph shows a deformed grain with heavily developed dislocation substructure being replaced by a new grain with no visible substructure. The boundaries of new grains such as these can often be shown to contain thin brine films. Scale bar, 0.1 mm. *e*, Lobate new grain (N) consuming heavily deformed old grains (O, P) in a recrystallized Asse sample deformed in non-dilatant conditions. Note that gas-bearing inclusions in the boundary between grains O and P have been overgrown by the new grain N to form a 'ghost' grain boundary. Scale bar, 0.1 mm. *f*, Reflection micrograph showing dynamic recrystallization in a rock salt sample from Avery Island deformed to ~40% strain at T = 70 °C and P = 20.4 MPa, with differential stresses in the range 10.3–20.4 MPa. Again, note how new grains with relatively little substructure are growing at the expense of the old, heavily substructured grains. Scale bar, 0.1 mm (sample courtesy of Dr F. D. Hansen). *g*, Reflected-light micrograph of deformed sample of artificially prepared salt (grain size ≃80 μm; see Fig. 4 for mechanical data). Note the fine, equiaxial, largely recrystallized grain structure. The dark spots are pores emerging at the surface of the polished section. Scale bar, 0.1 mm. Neutron diffraction methods have shown that this sample has no significant crystallographic preferred orientation (H. G. Brokmeier, personal communication). *h*, Fine-grained, recrystallized microstructure in a sample from the salt glacier described in ref. 14. Compare this with *a*, *d*, *f* and *g*. Scale bar, 0.1 mm. *i*, Scanning electron micrograph of a mechanically parted grain boundary in a sample of the Asse salt used as starting material in our experiments on natural material (see *b*). Note the tendency of the grain-boundary fluid inclusions to develop negative crystal forms. Scale bar, 0.04 mm. *j*, Scanning electron micrograph of grain surface in a recrystallized Asse sample (see *d*), one month after the end of deformation in non-dilatant conditions. Note the smoothness of these surfaces. Such grain boundaries can be shown to contain brine by disrupting them *in situ* under the transmission optical microscope. The brine film thickness has been shown to be <100 nm by measuring triple-junction void diameters from scanning electron micrographs. Scale bar, 0.1 mm. *k*, *l*, Scanning electron micrographs of grain surfaces in a recrystallized Asse sample, tested in identical conditions to the sample shown in *j*, but observed 1 yr after the test ended. Note the clear changes in morphology. It is inferred that the observed arrays of inclusions formed by the breakdown of originally subcontinuous brine films, presumably by means of a solution-precipitation process driven by surface energy. This also offers an explanation for the grain-boundary inclusion arrays seen in the Asse starting material (*b*, *i*). Scale bars, 0.02 mm.

2.5 MPa (and without added brine) show marked dilatancy during the first stage of deformation. During relaxation at strain rates of $>10^{-7}\ s^{-1}$, the differential stress is relatively insensitive to strain rate (Fig. 2). Microstructural analysis of the samples reveals densely developed deformation-band structures and crude subgrains (Fig. 3*c*). Grain boundaries are commonly dilated and fluid inclusions are disrupted (Fig. 3*c*). The overall behaviour agrees well with previous findings: work hardening is accompanied by some climb-controlled recovery[1,25].

(2) Samples deformed at 10 MPa confining pressure (with and without brine added) show little or no dilatancy during the first stage of the test. Furthermore, during stress relaxation there is a marked weakening below strain rates of $10^{-7}\ s^{-1}$ (Fig. 2). In these samples, the grain size is reduced to about half that of the starting material. Optical microscopy shows deformed grains being replaced by new, recrystallized grains (Fig. 3*d*, *e*, *f*). The samples are ~70% recrystallized. New grain boundaries contain subcontinuous brine films (Fig. 3*i*–*l*) which are <100 nm thick.

The most likely explanation of the type (2) behaviour is as follows. During the first stage of the experiments, brine present in grain-boundary voids (Fig. 3*b*) is transformed into subcontinuous brine films[26] which enhance grain boundary mobility. The grain boundaries then migrate, and the sample recrystallizes into a finer aggregate with lower dislocation density, and a network of brine-filled grain boundaries. The low values of the stress exponent n at strain rates below $10^{-7}\ s^{-1}$ (Fig. 2) suggest that the dominant deformation mechanism changes from dislocation to solution-transfer creep[8,18,19]. This is supported by the finding (Fig. 2) that the observed weakening behaviour agrees well with the following theoretical model for diffusion-controlled solution-transfer creep[8]:

$$\dot{\varepsilon} = \frac{ABC\sigma}{kTd^3} \qquad (1)$$

where $\dot{\varepsilon}$ is the strain rate, σ is the applied stress, A is a constant, B is a temperature-dependent diffusivity term, C is a grain-boundary structure parameter, k is the Boltzmann constant, T is the absolute temperature and d is the grain size.

At low strain rates, the samples showing type (2) behaviour are much weaker than expected from previous data (see Fig. 2), the weakening effect becoming more pronounced towards lower strain rates. The absence of this weakening in type (1) samples

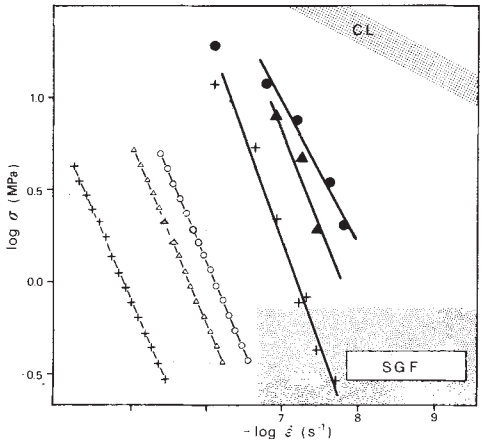


Fig. 4 Log σ plotted against log $\dot{\varepsilon}$ for three experiments on wet synthetic salt (solid lines), compared with the behaviour expected on the basis of previously existing creep laws[1,3,4,25] (stippled area, CL). The broken lines represent the behaviour predicted, for each of the three experiments reported, using the solution-precipitation creep model (eqn 1). Experiments: ●, $T = 20$ °C; $d = 0.2$ mm; $P = 21$ MPa; $P_f = 3$ MPa; ▲, $T = 70$ °C; $d = 0.2$ mm; $P = 21$ MPa; $P_f = 3$ MPa; +, $T = 20$ °C; $d = 0.08$ mm; $P = 5.1$ MPa; $p_f = 0.1$ MPa. Model: ○, $T = 20$ °C; $d = 0.2$ mm; △, $T = 70$ °C; +, $T = 20$ °C; $d = 0.08$ mm. Note the excellent agreement regarding stress sensitivity of strain rate ($n \approx 1$). The model was applied using $A = 44$ (ref. 18) and $c = 8 \times 10^{-8}$ m (ref. 8). The value of B was estimated from published data[8]. Finally, note that the experimental results reported in this diagram are almost directly relevant to the σ-$\dot{\varepsilon}$ conditions (stippled area, SGF) associated with the flow of wet salt glaciers in nature[13] ($d = 0.2$ mm).

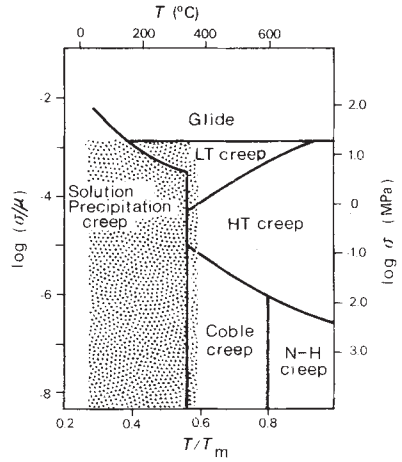


Fig. 5 Deformation mechanism map for halite (after ref. 28), modified to include solution-precipitation creep as described by the model given in the text (equation (1)). Values of the parameters A, B and C were chosen as for Figs 2 and 4. The map was constructed for a grain size of 1 mm and intergranular fluid pressure of 10 MPa. The extent of the solution-precipitation or solution-transfer creep field towards higher temperatures is assumed to be limited by the liquid-vapour phase transition. Stippling indicates the range of conditions relevant for halokinesis and radioactive waste repositories (the latter ranging up to only 200 °C). LT, low-temperature; HT, high-temperature; N-H, Nabarro-Herring.

(those samples deformed dry at pressure allowing dilatancy) can be explained by grain-boundary microcracking. This almost certainly leads to evaporation of the grain-boundary fluid.

In the second series of experiments we used artificial, fine-grained samples expected to exhibit accelerated solution-transfer creep effects (see equation (1)). These were prepared by dry compaction and annealing of fine-grained analytical-grade NaCl. The experiments were carried out in non-dilatant conditions using samples of 30 mm diameter and 60 mm length containing ~1% added brine[8]. A strain-rate stepping technique was used. The results indeed show the expected 'low n' behaviour (Fig. 4) and grain size dependence. Microstructural investigations show an equiaxial, recrystallized fabric, with little or no dislocation substructure (Fig. 3g), and no preferred lattice orientation. Grain boundaries were fluid-filled, with thicknesses of the same order as the recrystallized Asse samples. Most of the fluid was located in voids. Taken together, these observations provide evidence[20] that diffusive mass transfer was the main deformation mechanism, accompanied by grain-boundary sliding and migration. Comparison of the experimental data with the model of equation (1), applied using reasonable values of the parameters involved, gives good agreement for the stress exponent n (Fig. 4). Note also that the present experiments are almost directly relevant to the conditions found in salt glaciers (Figs 3h, 4), and show the weak behaviour reported in ref. 13.

Natural rock salt invariably contains small amounts of brine[16], part of which resides in grain-boundary fluid inclusions and films. Both sets of experiments reported here show that grain-boundary brine can have large effects on rheology and deformation mechanisms, by promoting dynamic recrystallization and solution transfer. In our opinion the weakening reported here has not been observed by previous workers on natural salt, either because the lowest strain rates reached were too high (Fig. 2), or because the water inherent in the samples was lost before testing. Note that most previous work has been performed in dilatant conditions. It is clear that in future long-term creep work, careful consideration should be given to recrystallization and grain-boundary fluid effects.

Keeping in mind that our natural (Asse) starting material is relatively dry, weakening by water should occur in a wide range of salt rocks during natural deformation[17,27]. A rough extrapolation of our data to the grain sizes, tempertures, and differential stresses[7] expected during natural salt deformation suggests that brine-related effects should become important at strain rates of $\sim 10^{-10}$ s^{-1} (just below those accessible in the laboratory), and we propose that most natural salt deformation occurs in the transition zone between the dislocation-creep and solution-transfer fields. We further conclude that the extrapolation of previous flow laws to predict long-term behaviour is questionable, unless the effects reported here are also taken into account. Construction of a deformation mechanism map for salt (Fig. 5), incorporating solution-precipitation creep (equation (1)), shows that brine effects can be expected throughout a very wide range of natural conditions, including those relevant to radioactive waste repositories and storage caverns.

We thank H. Gies, C. J. Talbot and F. D. Hansen for supplying samples, C. J. Peach and J. L. Liezenberg for discussions and assistance with experimental work, J. N. Boland for help with electron microscopy, H.-G. Brokmeier for measuring preferred lattice orientations, and the workshop staff of the HPT laboratory at Utrecht for technical assistance. The work was performed under contract with the Netherlands Ministry of Economic Affairs and the Commission of the European Communities (contract WAS-153-80-7-N) in connection with their research programmes on radioactive waste management and storage.

Received 25 June; accepted 30 September 1986.

1. Carter, N. L. & Hansen, F. D. *Tectonophysics* **92**, 275-333 (1983).
2. Jackson, M. & Talbot, C. J. *Bull. geol. Soc. Am.* **97**, 305-323 (1986).
3. Langer, M. In *Proc. 1st Conf. on the Mechanical Behaviour of Salt* (eds Hardy, H. R. & Langer, M.) 201-240 (Trans-Tech, Clausthel-Zellerfeld, 1984).
4. Herrmann, W., Wawersik, W. R. & Lauson, H. S. *Sandia natn. Lab. Rep.* No. SAND80-2712 (1980).
5. Wawersik., W. R., Herman, W., Montgomery, S. T. & Lauson, S. in *Proc. int. Symp. Rock Mechanics related to Caverns and Pressure Shafts* (ed. Wittke, W.) 1345-1356 (Balkema, Amsterdam, 1982).
6. Skrotzki, W. & Welch, P. *Tectonophysics* **99**, 47-61 (1983).
7. Carter, N. L., Hansen, F. D. & Senseny, P. E. *J. geophys. Res.* **87**, 9289-9300 (1982).
8. Spiers, C. J., Urai, J. L., Lister, G. S., Boland, J. N. & Zwart, H. J. in *Nuclear Science and Technology* (Office for Official Publs. of the European Communities, Luxembourg).

9. Urai, J. L., Spiers, C. J., Liezenberg, J. L., Franssen, R. C. & Peach, C. J. *Geol. Minb.* (in the press).
10. Müller, H. G. *Z. Phys.* **96**, 279-327 (1935).
11. Friedman, M., Dula, W. F., Gangi, A. F. & Gazonas, G. A. in *Proc. 1st Conf. on the Mechanical Behaviour of Salt* (eds Hardy, H. R. & Langer, M.) 19-36 (Trans-Tech, Clausthel-Zellerfeld, 1984).
12. Guillope, M. & Poirier, J. P. *J. geophys. Res.* **84**, 5557-5567 (1979).
13. Talbot, C. J. & Rogers, E. A. *Science* 208-395 (1980).
14. Talbot, C. J. *Spec. Pap. geol. Soc. Am.* **9**, 173-184 (1981).
15. Wenkert, D. D. *Geophys. Res. Lett.* **6**, 523-526 (1979).
16. Roedder, E. *A,. Miner.* **69**, 413-439 (1984).
17. Urai, J. L. *Tectonophysics* **96**, 125-157 (1983).
18. Raj, R. *J. geophys. Res.* **87**, 4731-4739 (1982).
19. Lehner, F. K. & Bataille, J. *Pure appl. Geophys.* **122**, 53-85 (1984).
20. Raj, R. & Ashby, M. F. *Metall. Trans.* **2**, 1113-1127 (1971).
21. Ruhenbeck, C. *Z. Physik.* **207**, 446-469 (1967).
22. Green, A. K. & Bauser, E. *J. appl. Phys.* **39**, 2769-2779 (1968).
23. Jockwer, N. thesis, Tech. Univ. Clausthal (1981).
24. Spiers, C. J., Urai, J. L. & Lister, G. S. in *Proc. 2nd Conf. Mechanical Behaviour of Salt*, Hannover (in the press).
25. Heard, H. C. *Am. geophys. Un. Geophys. Monogr.* **16**, 191-209 (1972).
26. Urai, J. L., Means, W. D. & Lister, G. S. *Am. geophys. Un. Geophys. Monogr.* **36**, 161-199 (1986).
27. Urai, J. L. *Tectonophysics* **120**, 285-317 (1985).
28. Albrecht, H. & Hunsche, U. *Fortschr. Miner.* **58**, 212-247 (1980).

Diagenesis of US Gulf Coast shales

K. Pye*, D. H. Krinsley† & J. H. Burton†

* Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK
† Department of Geology, Arizona State University, Tempe, Arizona 85287, USA

The burial diagenesis of mudstones is of major interest in petroleum geology[1–3]. Key questions which remain incompletely answered include the relative importance of temperature, time, changes in pore-water chemistry and organic maturation as driving mechanisms, and the relationship between diagenesis in mudstones and neighbouring sandstones[4–7]. We reconsider here the nature and timing of diagenesis in Gulf Coast Tertiary mudstones using data obtained by backscattered electron microscopy (BSEM) and energy-dispersive X-ray analysis (EDS) of cuttings from two Texas wells. Our observations confirm the mineralogical trends with depth reported by Hower *et al.*[1] and suggest new evidence concerning the timing and causes of the changes. The most important mineralogical changes occurred in the zone of organic matter decarboxylation (zone IV[8]). We show that the sequence of burial diagenetic events can be established quickly and reliably by BSEM examination of shale cuttings, and demonstrate the use of foram tests and their authigenic mineral infillings as indicators of diagenesis in soft, fine-grained shales.

The Tertiary mudstones and interbedded sandstones of the USA Gulf Coast represent one of the world's most intensively studied sedimentary sequences. Many fundamental concepts concerning the nature, timing and consequences of shale diagenesis were developed through work in this area[1,5,9–12]. Of particular interest has been the transformation of smectite in mixed-layer illite-smectite clays, involving an apparent increase in the proportion and ordering of illite layers with burial. It has been suggested that release of water during this transformation may be important in flushing out hydrocarbons from source rocks, in transporting ions for the formation and/or dissolution of cements in neighbouring reservoir sandstones, and in leading to over-pressuring and hydro-fracturing[3,9,13].

Several important questions have not been resolved in the classic work by Hower and his associates[1,11,14,15]. In particular, the sequence of mineralogical changes and the driving mechanism behind the diagenetic transformations were not positively identified. Decomposition of detrital potassium feldspar and mica during burial was suggested to provide the additional K^+ and Al^{3+} required for the transformation of smectite to illite[1], with release of Si^{4+}, Fe^{2+} and Mg^{2+} (reprecipitated as chlorite and quartz). It was suggested that substitution of Al^{3+} for Si^{4+} in the tetrahedral layers of the smectite was the principal factor allowing fixation of K^+ in the interlayer positions by increasing the layer negative charge. The reasons for the decomposition of K-feldspar and mica were not identified. Hower *et al.* also observed a gradual loss of calcite over the depth interval in which the clay transformations occurred (2,500–3,700 m), but no causal connection was suggested.

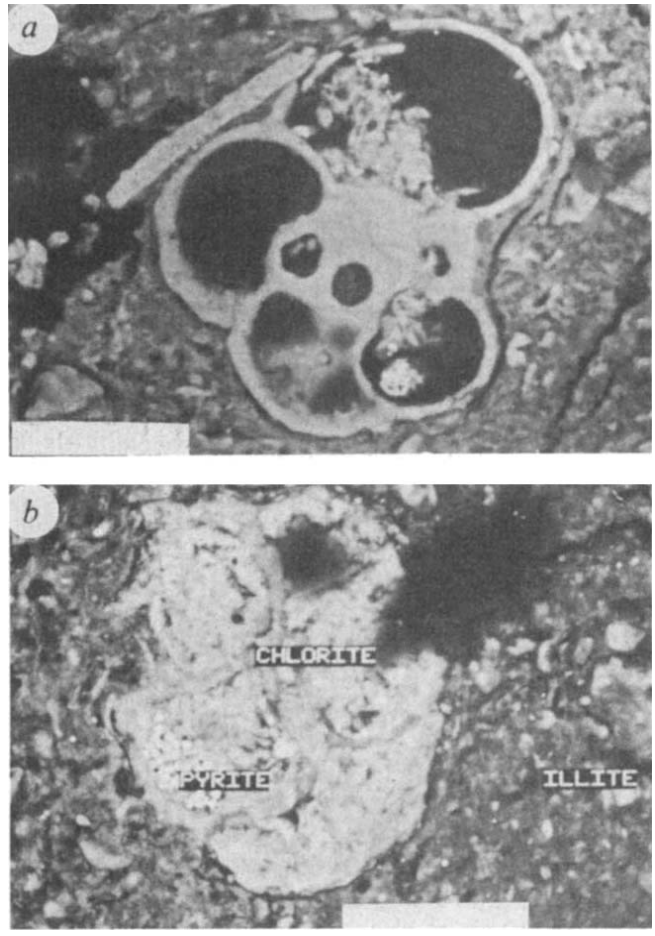


Fig. 1 BSEM micrographs of polished shale cuttings from C.W.R.U. Gulf Coast Well No. 6 showing: *a*, a calcite foram test partly infilled by authigenic chlorite (depth 2,680 m); and *b*, a foram completely infilled and replaced by chlorite (depth 3,415 m). The surrounding matrix consists mainly of illite-smectite and fine-grained quartz (scale bar, 50 μm).

De Segonzac[16] postulated that pressure is a major factor affecting clay transformations, but Perry and Hower[11] and Heling[17] concluded that temperature is more important. Other authors have stressed the role of pore-water chemistry, particularly K^+ availability[12,18]. Iron reduction[7,19] and organic matter maturation, leading to marked changes in shale porewater pH[5,6,20], have also been suggested to be important controls.

Previous Gulf Coast shale studies have relied mainly on 'bulk' methods of mineralogical and chemical analysis, particularly X-ray diffraction (XRD). No direct observations of textural relationships between different minerals with depth have been made, and diagenetic processes such as dissolution, precipitation and recrystalization have mainly been inferred from changes in relative mineral, elemental and isotopic abundance in different grain size fractions. We considered that useful information might be provided by examining polished shale cuttings using BSEM and EDS[21]. Cuttings from two south Texas wells were examined. The first sample suite spanned the depth interval 1,200–3,700 m in Case Western Reserve University Gulf Coast Well No. 6, previously studied by Hower *et al.*[1]. The second

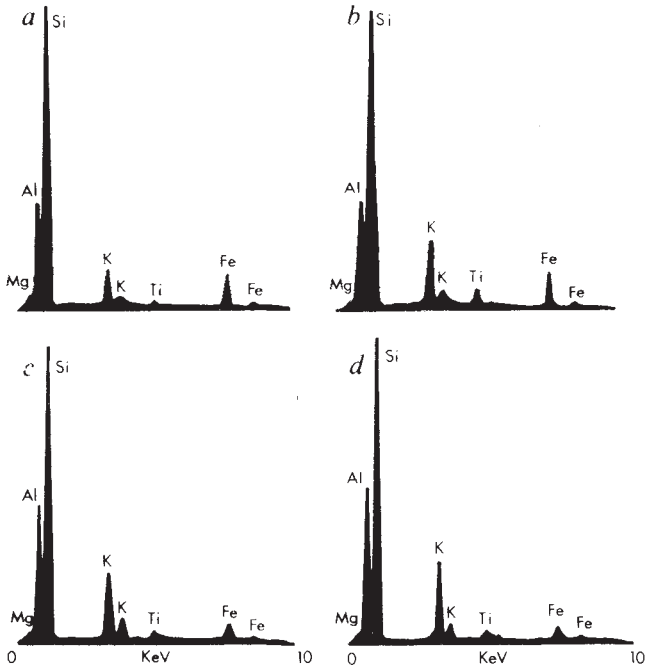


Fig. 2 Typical EDS spectra of illite-smectite aggregates showing observed changes in elemental peak heights with depth in Mobil Halls Bayou Wildcat No. 1. Samples are from depths of: *a*, 2,232 m; *b*, 3,650 m; *C*, 4,485 m; *d*, 5,530 m.

suite came from Mobil Halls Bayou Ranch No. 1 Wildcat, Galveston County. Both wells penetrate virtually the entire thickness of the Frio and Anahuac Formations which were deposited as a rapidly subsiding sediment wedge on the north side of the Gulf Coast Tertiary Basin[22,23]. The upper Frio consists of interbedded terrestrial and marginal marine sandstone and shale units underlain by >2,000 m of undercompacted slope and basinal mudstones. The Anahuac represents a transgressive, predominantly marine shale wedge which onlaps the Frio.

Shale cuttings from selected depth intervals were oriented and impregnated in epoxy for scanning electron microscope (SEM) examination transverse to the bedding. Due to their fine-grained, soft nature, many specimens were polished by hand to minimize surface damage. After carbon coating the specimens were examined in an ISI SEM equipped with a solid-state BSE detector and Kevex EDS analyser. Cuttings from each depth interval were also analysed by XRD.

Useful textural and chemical data were obtained from all but the finest-grained cuttings. Even in specimens composed largely of clay, clear indications of diagenetic reactions are provided by authigenic mineral aggregates within or replacing foram tests. In Gulf Coast Well No. 6, calcite foram tests are abundant above 2,750 m (Fig. 1*a*). The tests are partly filled by authigenic pyrite, kaolinite and chlorite, with some unoccupied pore space. Between 2,750 and 3,400 m the remaining voids are filled by authigenic chlorite and/or illite (Fig. 1*b*), while the tests themselves are dissolved. In some cases, the chambered structure of the tests is preserved by chlorite replacement, indicating that infilling of the intra-test voids occurred earlier than dissolution of the test walls. In the same depth interval, zoned authigenic carbonate rhombs, which were formed in the sulphate reduction and fermentation zones (zones II and III[8]), are also dissolved and replaced by iron-rich chlorite. These observations indicate a significant reduction in shale porewater *p*H in the temperature range 60–90 °C (see ref. 1 for down-hole temperature data).

Although the Frio mudstones contain many clay particles which cannot be resolved individually using BSEM, much of the illite-smectite clay (identified by XRD) occurs as aggregates up to 20 μm in size which may be compacted floccules. The aggregates are sufficiently mono-mineralic to allow semi-quantitative X-ray analysis of their chemical composition. The EDS spectra in Fig. 2 illustrate the general nature of the changes in composition of the illite-smectite aggregates with depth in the Mobil well. Similar trends were observed in C.W.R.U. Well No. 6. The illite-smectite above 2,500 m has a high Si/Al peak height ratio and a low K/Fe ratio, reflecting an abundance of smectite components. With increasing depth there is an increase in Al and K and a relative reduction in Si and Fe. This is in agreement with previous wet chemical data obtained from the fine-clay fractions[1] and corresponds with an increase in the proportion of illite components indicated by XRD[1]. In the Mobil well, the first change in composition of the illite-smectite indicated by EDS occurs at 3,620 m. At this depth, K abundance increases whereas Fe and Al abundance remains unchanged. With increasing depth there is a marked increase in Al and decrease in Fe while the K content remains almost constant. These observations suggest that the conversion of smectite to illite (if some form of solid-state transformation is occurring) is not dependent on tetrahedral substitution of Al^{3+} for Si^{4+}. However, it is possible that the creation of octahedral layer charge, due to the reduction of Fe^{3+}, during passage of the sediment through diagenetic zones II, III and IV, contributes to the fixation of interlayer K^+, as previously suggested by Eslinger *et al.*[24].

The infilling and replacement of some foram tests by illite indicates that this mineral is at least partly neoformed by direct precipitation from pore fluids.

K-feldspar was not present in large amounts in any of the cuttings studied, but bulk rock XRD analyses and SEM evidence indicated some reduction in abundance at slightly greater depth than carbonate dissolution, suggesting that carbonate may have acted as a temporary buffer which retarded dissolution of feldspar.

It has been suggested that thermal decarboxylation of kerogen in mudstones is the most probable mechanism whereby acid pore waters are generated at depth[20,25]. This may occur where production of CO_2 by decarboxylation exceeds iron reduction[7]. In the Gulf Coast, excess decarboxylation is favoured by a relatively high geothermal gradient and a rapid rate of burial. However, the Frio shales contain on average only 0.28% organic matter, and could not generate sufficient CO_2 to account for all of the secondary porosity observed in Frio sandstones plus the loss of carbonate originally present in the shales[5]. Aliphatic and other organic acid anions are also produced during decarboxylation[20,26], but a similar mass balance problem arises[5]. A further possibility is that, once initiated, the smectite-illite transformation generates residual acid, proton-donating water which may contribute to a lowering of *p*H[27,28].

From the textural and chemical data provided by BSE and EDS we conclude that the sequence of diagenetic changes was (1) reduction of *p*H due to excess decarboxylation during simultaneous iron reduction; (2) dissolution of carbonate; (3) dissolution of detrital feldspars and micas; (4) build-up of negative octahedral layer charge due to reduction of Fe^{3+} in mixed-layer clays; (4) fixation of interlayer K^+ and substitution of Al^{3+} for tetrahedral Si^{4+}, the additional K and Al being derived from dissolved detrital silicates; (5) precipitation of neoformed illite and chlorite in pores, and partial replacement of kaolinite by chlorite. Some chlorite and/or berthierine[29] was precipitated within foram tests before the onset of carbonate dissolution in zone IV.

The pattern of shale diagenesis found in the Gulf Coast differs from that found in some other sedimentary basins, and it may be unwise to regard it as a general model. The Gulf Coast sediments initially contained a large iron-bearing smectite and mixed-layer clay component, were buried relatively rapidly, and experienced high thermal gradients. Consequently they bear an important imprint from diagenetic zone IV. In contrast, many British Jurassic mudstones initially contained little smectite or

mixed-layer illite-smectite compared with kaolinite and degraded micas, and experienced low burial rates, with the result that the rocks bear a dominant mineralogical imprint from diagenetic zones II and III[30-32].

This research was supported by NATO grant RG 84/0039, the Petroleum Research Fund, the American Chemical Society, and the American Association of Petroleum Geologists.

Received 9 June; accepted 10 October 1986.

1. Hower, J., Eslinger, E. V., Hower, M. E. & Perry, E. A. *Bull. geol. Soc. Am.* **87**, 725-737 (1976).
2. Foscolos, A. E. & Powell, T. G. in *International Clay Conference 1978* (eds Mortland, M. M. & Farmer, V. C.) 262-270 (Elsevier, Amsterdam, 1978).
3. Bruce, C. H. *Am. Ass. petrol. Geol. Bull.* **68**, 673-683 (1984).
4. Boles, J. R. & Franks, S. G. *J. sedim. Petrol.* **49**, 55-70 (1979).
5. Lundgard, P. D., Land, L. S. & Galloway, W. E. *Geology* **12**, 399-402 (1984).
6. Surdam, R. C., Boese, S. W. & Crossey, L. J. *Am. Ass. petrol. Geol. Mem.* **37**, 127-149 (1984).
7. Curtis, C. D. *Phil. Trans. R. Soc.* A **315**, 91-105 (1985).
8. Curtis, C. D. *Geol. Soc. Lond. Spec. Publ.* 113-125 (1985).
9. Powers, M. C. *Am. Ass. petrol. Geol. Bull.* **51**, 1240-1254 (1967).
10. Burst, J. F. *Am. Ass. petrol. Geol. Bull.* **53**, 73-93 (1969).
11. Perry, E. A. & Hower, J. *Am. Ass. petrol. Geol. Bull.* **56**, 2013-1021 (1972).
12. Morton, J. P. *Bull. geol. Soc. Am.* **96**, 114-122 (1985).
13. Foster, W. R. & Custard, H. C. *Am. Ass. petrol. Geol. Bull.* **64**, 708 (1980).
14. Aronson, J. L. & Hower, J. *Bull. geol. Soc. Am.* **87**, 738-744 (1976).
15. Yeh, H-W. & Savin, S. M. *Bull. geol. Soc. Am.* **88**, 1321-1330 (1977).
16. De Segonzac, D. *Sedimentology* **15**, 281-346 (1970).
17. Heling, D. *Sedimentology* **21**, 463-472 (1974).
18. Roberson, H. E. & Lahann, R. W. *Clays Clay Miner.* **29**, 129-135 (1981).
19. Coleman, M. L. *Phil. Trans. R. Soc.* A **315**, 39-56 (1985).
20. Curtis, C. D. *Am. Ass. petrol. Geol. Bull.* **67**, 380-383 (1983).
21. Pye, K. & Krinsley, D. H. *J. sedim. Petrol.* **54**, 877-888 (1984).
22. Loucks, R. G., Bebout, D. G. & Galloway, W. E. *Trans. Gulf Coast. Ass. geol. Soc.* **27**, 109-120 (1977).
23. Galloway, W. E., Hobday, D. K. & Magara, K. *Bur. Econ. Geol. Univ. Texas, Austin, Rep. Investign* **122**, 1-78 (1982).
24. Eslinger, E., Highsmith, P., Albers, D. & DeMayo, B. *Clays Clay Miner.* **27**, 327-338 (1979).
25. Schmidt, V. & McDonald, D. A. *Soc. Econ. Paleontol. Miner. Spec. Publ.* **26**, 175-207 (1979).
26. Carothers, W. W. & Kharaka, Y. K. *Am. Ass. petrol Geol. Bull.* **62**, 2441-1453 (1978).
27. Mortland, M. M. *Trans. 9th int. Cong. Soil Sci.* **1**, 691-699 (1968).
28. Johns, W. D. & Shimoyama, A. *Am. Ass. petrol. Geol. Bull.* **56**, 2160-2167 (1972).
29. Ahn, J. H. and Peacor, D. R. *Clays Clay Miner.* **33**, 228-236 (1985).
30. Irwin, H., Curtis, C. D. & Coleman, M. L. *Nature* **269**, 209-213 (1977).
31. Morris, K. A. *J. geol. Soc. Lond.* **137**, 157-175 (1980).
32. Pye, K. & Krinsley, D. H. *Geol. Mag.* **123**, 191-203 (1986).

Do continental shelves export organic matter?

G. T. Rowe*, S. Smith*, P. Falkowski*, T. Whitledge*, R. Theroux†, W. Phoel‡ & H. Ducklow§

* Oceanographic Sciences Division, Department of Applied Science, Brookhaven National Laboratory, Upton, New York 11973, USA
† Northeast Fisheries Center, NMFS/NOAA Woods Hole, Massachusetts 02543, USA
‡ Northeast Fisheries Center, NMFS/NOAA Sandy Hook, New Jersey 07732, USA
§ University of Maryland, Center for Environmental Studies, Cambridge, Maryland 21613, USA

It has been suggested that biological production and consumption of organic matter is not balanced in coastal marine ecosystems[1,2]. If, as suggested, 90% of the phytoplankton produced during the spring bloom period were exported, excess organic carbon would be sequestered on the continental slope below the permanent thermocline. Here we summarize the shelf-edge exchange processes (SEEP) experiment, designed to test the export hypothesis. The absence of a positive imbalance in the organic carbon budget, reinforced by modest sediment deposition and biomass on the continental slope, led us to reject the concept. Only a small fraction of continental shelf phytodetritus is exported; that not consumed in the spring is for the most part used on the continental shelf during the ensuing stratified season. The original hypothesis failed to recognize the contribution of pelagic microbial consumption and the lag in coupling between seasonal production and consumption processes.

Table 1 Moored sediment trap (MST) flux

	Date	Mass flux ($mg\ m^{-2}\ d^{-1}$)	C ($mg^{-2}\ d^{-1}$)	N ($mg\ m^{-2}\ d^{-1}$)
MST 1	25 Feb.-5 March 1984			
70 m.o.b.*		1,240 ± 198	54 ± 6	6 ± 1
10 m.o.b.		5,450 ± 467	199 ± 44	20 ± 7
MST 2	5-31 March 1984			
70 m.o.b.		1,810 ± 239	110 ± 28	12 ± 5
10 m.o.b.		3,815 ± 418	159 ± 18	19 ± 4
MST 3	4-7 April 1984			
70 m.o.b.		2,298 ± 240	140 ± 12	6 ± 4
10 m.o.b.		2,748 ± 624	107 ± 16	1 ± 1

Data from SEEP I (150 m, west line locality W12 shown on Fig. 1). Values are mean ± s.d.
* m.o.b., metres off bottom.

The data we present come from a restricted area of the continental shelf of the eastern United States south of New England, which extends from longitudes 70°30′ to 72°30′ W., and includes the area between the coastline and the outer edge of the continental shelf from Martha's Vineyard south to Long Island (Fig. 1). Two transects of moorings were instrumented with current meters, transmissometers and internally recording fluorometers[3,4] between February and April 1984. Sediment traps[5] were moored at locality W12 (Fig. 1) during this period, and at the deeper moorings on the east line[6] throughout the remainder of the year. Estimates of standing stocks and a variety of biological rate measurements were made during 24-48 h periods in shelf and slope water on both lines (Fig. 1) between late February and mid-April 1984. Measurements over similar periods were made during July 1983. On the basis of the original hypothesis these data should have shown that annual net primary production appreciably exceeded consumption on the continental shelf; that biogenic particles or plant pigments had a net offshore flux; and that intensified organic carbon deposition was reflected in the sediments or the biota on the continental slope.

The main aim of our work was to study the shelf carbon budget and to search for an imbalance betweeen primary production and heterotrophic use within the study area (Fig. 1). Primary productivity was estimated by measuring ^{14}C-labelled assimilation during incubations aboard ship[7], and the consumption of phytoplankton carbon was estimated by summing respiration and secondary production of macrozooplankton[8-10], the benthic community[11,12] and pelagic bacteria[13,14]. Microzooplankton were not included. Zooplankton grazing was estimated from gut clearance rates[10] and known relationships between feeding rates, body size and food concentration[9]. Bacterial respiration and production were based on labelled thymidine incorporation[13] and NH_4 regeneration rate measurements using ^{15}N-labelled substrates[14]. Use of organic matter by the benthic biota was assumed to be by respiration and secondary production of the macrofauna[15]. Respiration, the sum of anaerobic microbial metabolism and oxygen use, was estimated from oxygen use *in situ* in benthic chambers placed on the bottom by scuba divers and by ship-board incubations of the sediment cores[12]. Sulphate reduction, nitrate reduction and organic carbon accumulation in shelf sediments may also contribute to carbon fluxes[12], but were not included here. Although total benthic standing stocks were measured[12], we include here only the secondary production of the macrofauna based on established production to biomass ratios[15,16] and measured biomass of the macrofauna[17,18]. Secondary production of the meiofauna and benthic bacteria were not included.

The critical issue is the comparison of phytoplankton production rate with the total rate of consumption of organic matter.

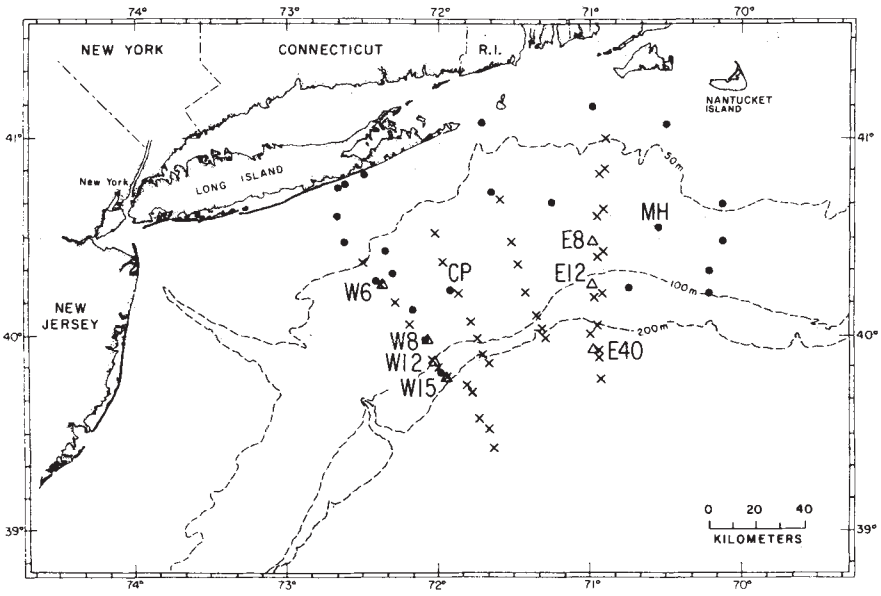
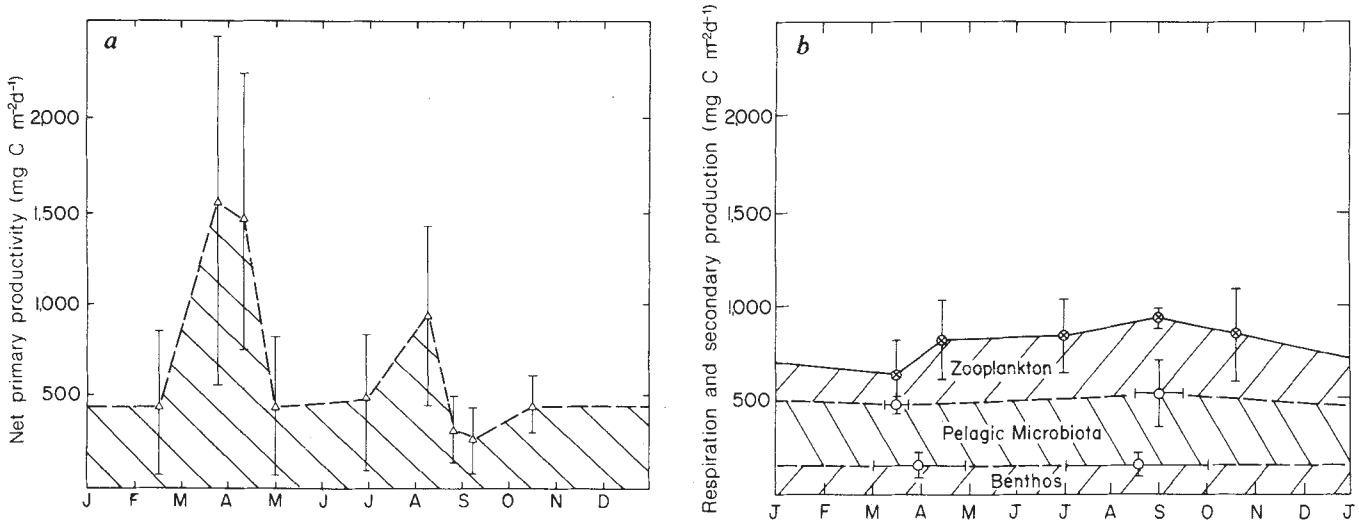


Fig. 1 Location of SEEP experiment, east and west lines. Time series were conducted at the triangles. Sediment traps were located at W12. ●, Locations of sediment oxygen demand measurements; ×, locations of core or grab samples.



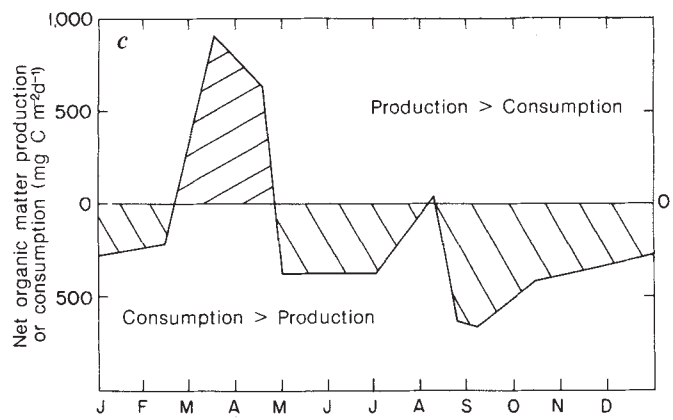


Fig. 2 *a*, The ^{14}C incorporation primary production rates[7] and *b*, sum of heterotrophic consumption rates in the SEEP area. All benthic rates summed and averaged for the year include oxygen demand and secondary production of the macrofauna[11,12], the pelagic bacterial rates of respiration and production for the winter/spring[13] and the late summer[14] periods and macrozooplankton assimilation[10]. Data and methods from refs 3-19. *c*, Consumption rates shown in *b*, subtracted from the primary production rate shown in *a*. Error bars, standard deviations of individual data sets.

We have plotted the two separately over the course of a year (Fig. 2*a*, *b*). The seasonal contrast is obvious on subtraction of integrated consumption from production (Fig. 2*c*). This suggests that carbon fixation was predominant in the spring but that total remineralization and secondary production were the most significant processes during the rest of the year. The results presented in Fig. 2*c* provide no evidence to support the hypothesis that significant quantities of organic matter are available for export. An annual manifestation of the intense heterotrophic activity following the bloom is the recurrent hypoxia experienced in the mid-Atlantic bight every year during stratified periods[19].

The budgeting approach undoubtedly suffers from some inaccuracies due to imprecision in each component. The ^{14}C method may underestimate primary productivity (Fig. 2*a*) if there is a loss of smaller photosynthetic picoplankton[20]. Consumption (Fig. 2*b*), on the other hand, does not include microzooplankton. But as the microzooplankton consume the picoplankton as well as bacteria, the two omissions could to some degree cancel each

other out[21]. Sediment anaerobic respiration that is not incorporated into the measured total sediment oxygen demand (a distinct possibility if sulphide forms pyrite or is otherwise immobilized[12]) has also not been taken into account. In spite of the limitations of our data set, the negative imbalance in the carbon budget (Fig. 2*c*) cannot be considered justification for accepting the export hypothesis.

Fluxes of organic carbon into sediment placed on the west line (Fig. 1) also provided little evidence for accepting the hypothesis. One mooring array of two traps was set for three different periods between 25 February and 7 April 1984, within continental slope water but underlying continental shelf water (locality W12 on Fig. 1). Extensive particle export by sinking through or mixing across the shelf-slope front should be reflected in the traps, but the average rate of organic carbon flux to all traps (Table 1) was only 128 mg C m^{-2} d^{-1} ($\sigma = \pm 45$), or about 10% of the integrated euphotic zone primary productivity ($\sim$1.5 g C m^{-2} d^{-1}, $\sigma = \pm 0.76$) at that time[7]. This small fraction of the overlying net primary productivity was about what would be expected from direct vertical transport[22,23] and was below what we would expect if 90% of the net primary production during the spring bloom was exported to this depth[1,2]. If only 50% of the net productivity was exported from a 100 km-wide continental shelf, and concentrated on a 30 km-wide continental slope, we would expect the organic carbon particle flux to be $\sim$1–2 g C m^{-2} d^{-1}, which is ten times greater than that sampled by the traps (Table 1).

Concurrent investigations of the transport of plant pigments and the characteristics of the sediments and biota of the continental slope support the above conclusions. Although fluxes into sediment traps lower on the slope were greater near bottom than in the water column, the flux rates were far less than those we measured (Table 1) adjacent to the shelf-slope front[6]. The moored fluorometers documented rapid sinking and extensive along-shore advection of biogenic particles. Although some equivocal evidence suggested a net offshore transport may have occurred on the shelf, there was no evidence that appreciable material crossed the shelf-slope front[4]. Cross-front exchanges of water between distinct shelf and slope water masses were very limited during the spring[24], although turbulent mixing may have been accentuated at the bottom by internal waves[25]. An exchange of water of $\sim$20% across the front as it moves from Cape Cod to Cape Hatteras, based on salt and nitrate budgets[7], could be accounted for by this bottom mixed-layer exchange[25].

The logic behind extensive organic carbon export has been questioned before[26,27], but our data graphically illustrate that seasonal coupling and pelagic microbial activity[28] make the idea untenable. Our results do not support the suggestion that an appreciable fraction of the global missing carbon is on continental slopes[1], but they do not exclude the possibility that most of the carbon on continental slopes was due to shelf export, albeit at a relatively slow rate. Although our rejection of the hypothesis applies to the north-west Atlantic SEEP area in particular, it is important to ask how our conclusions might apply to other shelf ecosystems. At higher latitudes, for example, colder water may depress microheterotrophic activity[29]. The export and fate of carbon from narrow shelves where primary production is enhanced by external sources of 'new' nutrients[2,30,31] remains to be assessed.

This work was part of continuing studies of coastal zone ecosystems by the Office of Health and Environmental Research, US Department of Energy (contract DE-ACO2-76CH00016), and the Northeast Fisheries Center, NMFS/NOAA. We thank our collaborators in the SEEP experiments, as well as the officers and crew of the RV *Oceanus*.

Received 22 May; accepted 25 September 1986.

1. Walsh, J. J., Rowe, G. T., Iverson, R. L. & McRoy, C. P. *Nature* **291**, 196–201 (1981).
2. Malone, T., Hopkins, T. S., Falkowski, P. C. & Whitledge, T. *Cont. Shelf Res.* **1**, 305–337 (1983).
3. Whitledge, T. E. & Wirick, C. D. *Deep Sea Res.* **30**, 29–309 (1983).
4. Walsh, J. J., Wirick, C., Pietrafesa, L. & Whitledge, T. & Hoge, F. *Cont. Shelf Res.* (in the press).
5. Rowe, G. & Gardner, W. *J. mar. Res.* **37**, 581–600 (1979).
6. Biscaye, P. & Anderson, R. *Cont. Shelf Res.* (submitted).
7. Falkowski, P., Rowe, G., Smith, S., Whitledge, T. & Flagg, C. *Cont. Shelf Res.* (in the press).
8. Dagg, M. J. & Grill, D. *Limnol. Oceanogr.* **25**, 597–609 (1980).
9. Dagg, M. J. & Turner, J. F. *Can. J. Fish. aquat. Sci.* **39**, 979–990 (1982).
10. Smith, S. & Lane, P. V. Z. *Cont. Shelf Res.* (in the press).
11. Florek, R. & Rowe, G. *Int. Rev. ges. Hydrobiol.* **68**, 73–112 (1983).
12. Rowe, G. T. *et al. Cont. Shelf Res.* (in the press).
13. Ducklow, H. & Kirchman, D. L. *J. Plank. Res.* **5**, 333–355 (1983).
14. Harrison, W. G., Douglas, D., Falkowski, P., Rowe, G. & Vidal, J. *J. Plank. Res.* **5**, 539–556 (1983).
15. Warwick, R. M., George, C. & Davies, J. R. *Est. Coast. mar. Sci.* **7**, 2–5 (1978).
16. Banse, K. & Mosher, S. *Ecol. Monogr.* **50**, 355–379 (1980).
17. Wigley, R. L. & Theroux, R. B. *Prof. Pap. U.S. geol. Surv.* **529**-N (1981).
18. Collie, J. S. *Mar. Ecol. Prog. Ser.* **22**, 229–238 (1985).
19. Falkowski, P. G., Hopkins, T. S. & Walsh, J. J. *J. mar. Res.* **38**, 479–506 (1980).
20. Glover, H. E., Campbell, L. & Prezlin, B. B. *Mar. Biol.* **91**, 193–203 (1986).
21. Goldman, J. C. & Caron, P. A. *Deep Sea Res.* **32**, 899–915 (1985).
22. Suess, E. *Nature* **288**, 260–263 (1980).
23. Lorenzen, C. J., Welschmeyer, N. & Copping, A. E. *Deep Sea Res.* **30**, 639–643 (1983).
24. Houghton, R. W., Aikman, F. & Ou, H. W. *Cont. Shelf Res.* (in the press).
25. Flagg, C. N. *Cont. Shelf Res.* (in the press).
26. Peng, T.-H. & Broecker, W. *J. geophys. Res.* **89**, 8170–8180 (1984).
27. Pace, M. L., Glasser, J. & Pomeroy, L. R. *Mar. Biol.* **82**, 47–63 (1984).
28. Ducklow, H. W., Purdie, D. A., Williams, P. LeB. & Davies, J. M. *Science* **232**, 865–867 (1986).
29. Pomeroy, L. R. & Deibel, D. *Science* **233**, 359–361 (1986).
30. Walsh, J. J. *Nature* **290**, 300–304 (1981).
31. Rowe, G. & Howarth, R. *Deep Sea Res.* **32**, 43–55 (1985).

Identification of novel widely distributed sedimentary acyclic sesterterpenoids

J. N. Robson & S. J. Rowland

Department of Environmental Sciences, Plymouth Polytechnic, Drake Circus, Plymouth, Devon PL4 8AA, UK

Isoprenoids, the general class of natural products biosynthesized from isoprene units, are abundant in nature. Many isoprenoids are also found in sediments, sedimentary rocks and crude oils where they have proved useful as 'chemical fossils' of biological activity and as indicators of geothermal stress[1]. However, acyclic isoprenoids with 25 carbon atoms—so-called sesterterpenoids—have been reported only rarely in the biosphere and the geosphere. Indeed, the only acyclic sesterterpenoids of confirmed structure reported in sediments are of pentamethyleicosane structural isomers (Fig. 1, carbon skeleton I). For example, 2,6,10,14,18-pentamethyleicosane has been proposed as a marker for halophilic bacteria and the 2,6,10,15,19-isomer (I) as an indicator of methanogens[2]. We have now identified a group of novel sesterterpenoids that are major components of the hydrocarbons of recent freshwater, marine and hypersaline sediments from many different parts of the globe. Because they have unusual structures, these sesterterpenes show promise as a new class of biological marker compounds.

Over the past 10 years many workers (Table 1) have reported the presence, in recent sediments, of a group of C_{25} hydrocarbons of unknown structure. The compounds are abundant and are very widely distributed; for example, they are the major hydrocarbons present in sediments from a New England salt marsh, Narragansett Bay (USA)[3]; North-east Gulf of Mexico[4]; the Peru upwelling region[5]; Shark Bay, Western Australia[6] and the Ebro River, Spain[7]. The same compounds have been detected in aquatic filtering organisms and in sediment traps suspended in the water column[7–10] and recently a C_{25} diene has been detected in field specimens of the green alga, *Enteromorpha*[11]. Numerous structures have been proposed for these C_{25} compounds (see Fig. 1, skeletons II–V)[7,11,12] although no structure has been confirmed. The electron impact mass spectrum and gas chomatographic Kovats Index (KI) of the parent C_{25} alkane[12,13] and ozonolysis of a C_{25} monoene[6] suggest structure II as the most likely skeleton. This is supported by synthesis of a co-occurring

Table 1 Occurrence of alkanes and alkenes with skeletons II, VII and VIII in sediments*

Locations	Age/environment	Carbon skeleton II	VII	VIII	Refs
Kiel Bight, FRG	Recent/marine	√		√	17, 18
Scotian Shelf, Nova Scotia	Recent/marine	√			19
Shark Bay, Western Australia	Recent/marine/hypersaline	√	√		6
Alaskan Outer Continental Shelf	Recent/marine	√		√	20
Peru Continental Shelf	Recent/marine	√			5, 21
Alfacs Bay, Ebro River, Spain	Recent/marine	√	√		7, 22
Dabob Bay, Washington State, USA	Recent/land-locked marine	√		√	9
Northeast Gulf of Mexico	Recent/marine	√	√		4
Southern California, USA	Recent/marine	√			10
Rhode Island Sound, USA	Recent/estuarine	√			8
Buzzards Bay, Massachusetts, USA	Recent/estuarine	√			23
Narragansett Bay, Rhode Island, USA	Recent/estuarine	√		√	3, 15
Puget Sound, Washington State, USA	Recent/land-locked marine	√	√	√	12, 13
Sandyhaven, Dyfed, Wales	Recent/intertidal	√	√		11
Grasmere, Cumbria, UK	Recent/freshwater lake		√		14
Camloch, Scotland	Recent/freshwater lake		√		24
Loch Clair, Scotland	Recent/freshwater lake		√		24
Upton Broad, Norfolk, UK	Recent/freshwater lake		√		24
Rostherne Mere, Cheshire, UK	Recent/freshwater lake	√	√		11
Cariaco Trench, Venezuela	Pleistocene/marine		√		25
Sabon-Gida, Nigeria	Pleistocene/lacustrine	√	√		25
Loch Ewe, Scotland	Recent/marine loch		√		25
Esthwaite Water, Cumbria, UK	Recent/freshwater lake		√		26
Pettaquamscutt River, Rhode Island, USA	Recent/estuarine	√		√	27
Sullom Voe, Shetland Islands, UK	Recent/intertidal	√	√		28

* The same compounds have been found in marine filtering organisms[8], an alga[11] and certain crude oils[25].

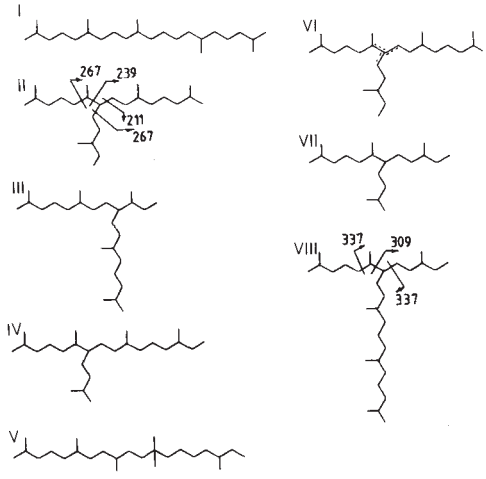


Fig. 1 Isoprenoids, carbon skeletons I–VIII.

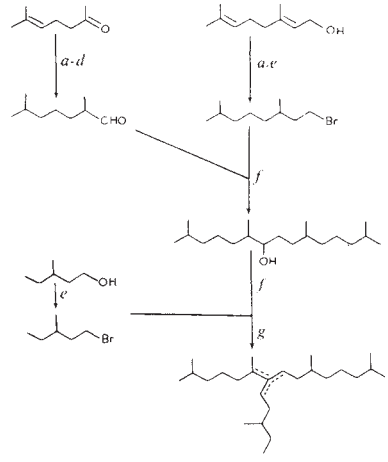


Fig. 2 Synthetic scheme for alkane II and alkenes VI. All products characterized by ^{1}H NMR mass and infrared spectroscopy. The second Grignard reaction (f) was repeated in improved yield (82%) by use of $CeCl_3$/Mg (ref. 29). *a*, H_2,PtO_2,hexane; *b*, $BrCH_2Br$,Mg; *c*, B_2H_6,THF,H_2O_2, OH *d*, $(COCl)_2$,DMSO; *e*, HBr/H_2SO_4; *f*, Mg,THF; *g*, $POCl_3$,Pyr.

C_{20} alkane (assumed to be pseudohomologous with the C_{25} alkane)[14]. Synthesis of V has shown this structure to be incorrect[12] and attempts to isolate the C_{25} alkenes by argentatious thin layer chromatography in quantities large enough for spectral characterization have been unsuccessful[15].

A mixture of five monoenes related to alkane II was synthesized in good yield from readily available starting materials by the scheme outlined in Fig. 2. Catalytic hydrogenation afforded the alkane. The electron impact mass spectrum of the synthetic alkane II is shown in Fig. 3. The spectrum, which is virtually identical to that of the sedimentary alkane[12] contains small but characteristic ion doublets at m/z 210,211; 238,239 and 266; 267 assigned to fragmentation about the C7–C8, C6–C7 and C5–C6 (and C7–C1′) bonds (structure II). The KI value for both synthetic and sedimentary alkane is 2,109 (DB-1, 40–290 °C at 4 °C min^{-1}). These data confirm that the C_{25} alkenes and alkane identified in many previous studies (Table 1 and ref. 11) possess the 2,6,10,14-tetramethyl-7-(3-methylpentyl) pentadecane skeleton (structure II). Mass spectra of the synthetic monoenes (Fig. 1, structure VI) were characterized by even mass ions at m/z 154,196,210 (weak), 224,266,280 and 350($M^{\cdot}+$) but none was identical to the spectra of any of the sedimentary monoenes. This is further indirect evidence that the sedimentary alkenes may indeed contain a methylene double bond as suggested previously from ozonolysis[6] and nuclear magnetic resonance (NMR)[8]. In sediments from Puget Sound[12] and Dabob Bay[9], Washington State (USA), the C_{25} hydrocarbons co-occurred with the C_{20} alkane (now known to be structure VII in Fig. 1[11]) and a C_{30} alkene of unknown structure but which from retention index measurements was considered to be pseudohomologous with both structures VII and II when hydrogenated. Given the structural relationship between VII and II demonstrated in the present study, it is reasonable to speculate that the C_{30} alkene is VIII (Fig. 1). This is consistent with the mass spectrum in that fragmentation about the C6–C7, C5–C6

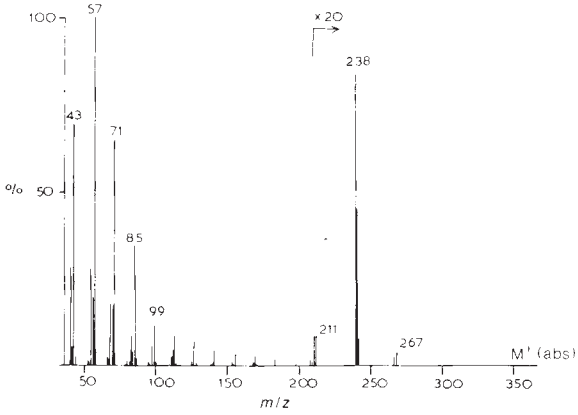


Fig. 3 Electron impact mass spectrum of synthetic II. Gas chromatography mass spectrometry conditions: Kratos MS25, 30 m × 0.3 mm CPSil 5 fused silica column, 40-300 °C at 4 °C min^{-1}. He carrier, 40 eV.

and C7-C1 bonds would produce the ions observed at m/z 308,309; 336,337 (structure VIII). Thus, the abundant sesterterpenes appear to be the major and most commonly occurring components of a new class of terpenoids comprising $C_{20,25,30}$ and possibly other[6], hydrocarbons.

These highly branched hydrocarbons have already proved useful as environmental indicators. For instance, variations in their abundance in sediments have been successfully used to designate distinct chemogeographic regions. Dunlop and Jefferies[6] were able to classify oceanic and hypersaline sedimentary basins in Western Australia on this basis and Requego and Quinn[15] were able to show that the geographical variations of some of the alkenes reflected the distribution of marine organic matter in the estuary of Narragansett Bay, USA. Now that the structures of the C_{20} and C_{25} hydrocarbons have been established and one possible source organism proposed[11], future studies may be based on a firmer understanding. Further synthesis and incubation of synthetic ^{14}C-labelled analogues should help to elucidate the longer term sedimentary fate of these new biological markers.

Finally, the unusual structures of these molecules suggest that they may have a specific role in the lipid biochemistry of the source organisms. This has proved to be the case for other unusual isoprenoid lipids. For instance biphytanyl isoprenoid ethers act as rigidifiers in the cell membranes of archaebacteria[16] and are therefore of interest to biologists, biochemists and geochemists. Further study of the present multi-branched lipids may benefit from a similar multidisciplinary approach.

Received 1 September; accepted 3 October 1986.

1. Tissot, B. P. & Welte, D. H. *Petroleum Formation and Occurrence* (Springer, Berlin, 1984).
2. Brassell, S. C., Wardroper, A. M. K., Thompson, I. D., Maxwell, J. R. & Eglinton, G. *Nature* **290**, 693-696 (1981).
3. Requejo, A. G. & Quinn, J. G. *Estuar. Coast. Shelf Sci.* **20**, 281-297 (1985).
4. Gearing, P., Gearing, J. N., Lytle, T. F. & Lytle, J. S. *Geochim. cosmochim. Acta* **40**, 1005-1017 (1976).
5. Volkman, J. K., Farrington, J. W., Gagosian, R. B. & Wakeham, S. G. in *Advances in Organic Geochemistry, 1981* (eds Bjoroy, M. *et al.*) 228-240 (Wiley, New York, 1983).
6. Dunlop, R. W. & Jefferies, P. R. *Org. Geochem.* **8**, 313-320 (1985).
7. Albaiges, J. *et al. Org. Geochem.* **6**, 237-248 (1984).
8. Boehm, P. D. & Quinn, J. G. *Estuar. Coast. mar. Sci.* **6**, 471-494 (1978).
9. Prahl, F. G., Bennett, J. T. & Carpenter, R. *Geochim. cosmochim. Acta* **44**, 1967-1976 (1980).
10. Crisp, P. T., Brenner, S., Venkateson, M. I., Ruth, E. & Kaplan, I. R. *Geochim. cosmochim. Acta* **43**, 1791-1801 (1979).
11. Rowland, S. J., Yon, D. A., Lewis, C. A. & Maxwell, J. R. *Org. Geochem.* **8**, 207-213 (1985).
12. Barrick, R. C. & Hedges, J. I. *Geochim. cosmochim. Acta* **45**, 381-392 (1981).
13. Barrick, R. C., Hedges, J. I. & Peterson, M. L. *Geochim. cosmochim. Acta* **44**, 1349-1362 (1980).
14. Yon, D. A., Ryback, G. & Maxwell, J. R. *Tetrahedron Lett.* **23**, 2143-2146 (1982).
15. Requejo, A. G. & Quinn, J. G. *Geochim. cosmochim. Acta* **47**, 1075-1090 (1983).
16. De Rosa, M., De Rosa, S., Gambacorta, A., Minale, L. & Bu'Lock, J. D. *Phytochemistry* **16**, 1961-1965 (1977).
17. Osterroht, C. & Petrick, G. *Mar. Chem.* **11**, 55-70 (1982).
18. Ehrhardt, M. & Heinemann, J. *Environ. Pollut.* **9**, 263-282 (1975).
19. Keizer, P. D., Dale, J. & Gordon, D. C. *Geochim. cosmochim. Acta* **42**, 165-172 (1978).
20. Venkatesan, M. I. & Kaplan, I. R. *Geochim. cosmochim. Acta* **46**, 2135-2149 (1982).
21. Smith, D. J., Eglinton, G. & Morris, R. J. *Geochim. cosmochim. Acta* **47**, 2225-2232 (1983).
22. Bayona, J. M. *et al. J. High Res. Chromatography chromatogr. Commun.* 605-611 (1983).
23. Farrington, J. W., Frew, N. M., Gschwend, P. M. & Tripp, B. W. *Estuar. Coastal Mar. Sci.* **5**, 793-808 (1977).
24. Cranwell, P. A. *Prog. Lipid Res.* **21**, 271-308 (1983).
25. Yon, D. A. thesis, Univ. Bristol (1981).
26. Brooks, P. W. *et al. Chem. Geol.* **20**, 189-204 (1977).
27. Requejo, A. G., Quinn, J. G., Gearing, J. N. & Gearing, P. J. *Org. Geochem.* **7**, 1-10 (1984).
28. Jones, D. M. thesis, Univ. Newcastle-on-Tyne (1986).
29. Imamoto, T., Takiyama, N. & Nakamura, K. *Tetrahedron Lett.* **26**, 4763-4766 (1985).

Ostrich ancestors found in the Northern Hemisphere suggest new hypothesis of ratite origins

Peter Houde

Division of Birds, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20560, USA

Modern ratites (ostriches, rheas, cassowaries, emus, and kiwis) are flightless birds which have a palatal structure termed 'palaeognathous'[1] and are found on daughter-landmasses of the Mesozoic supercontinent Gondwanaland. It has been suggested[2-4] that a single flightless ancestor, widely distributed in Gondwanaland, gave rise to the various lineages of ratite birds. The temporal calibration of the DNA molecular clock is primarily based on the divergence of ratites, and depends on the validity of this hypothesis. Newly studied fossils suggest that the ancestors of ostriches are instead among a group of North American and European birds, the '*Lithornis*-cohort', that had the potential of flight and from which the kiwis may have arisen separately.

The *Lithornis*-cohort of fossils is a group of hen-sized, palaeognathous carinates which had the potential of flight and are found in Palaeocene and Eocene deposits in North America and Europe[5]. The group includes *Lithornis*, *Paracathartes* and a genus I will refer to here as the 'Green River palaeognath' ('GRP'). They are the closest relatives of the ratite birds known to have had the capacity for flight, closer than the palaeo-

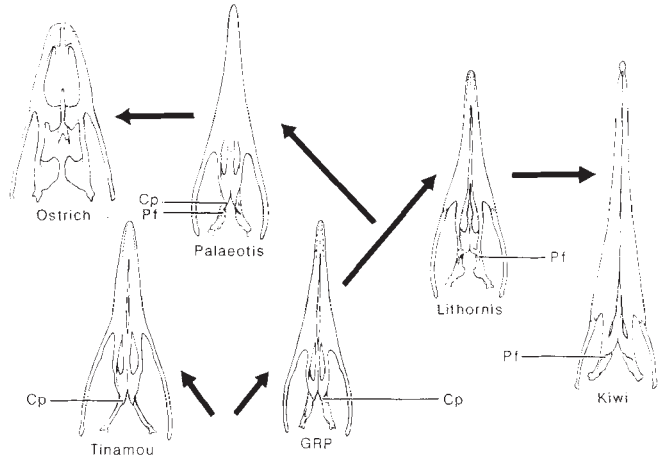


Fig. 1 Grades of evolution in the palatal complex of bones in the fossil and extant palaeognathous birds: tinamou (*Tinamus*, drawn from Smithsonian specimen USNM 345738): 'Green River palaeognath' (drawn from USNM 336103); *Lithornis* (drawn from USNM 391983; agrees with *Paracathartes*, USNM 391984); *Palaeotis* (drawn from HLMD Me 7530); kiwi (drawn from USNM 19025) and ostrich (drawn from USNM 224852), not to scale. The branching pattern does not necessarily represent the correct phylogenetic relationships of the fossil taxa but the monophyly of the ostrich and kiwi and the sister group relationship of tinamous is supported by molecular data[4]. Cp, caudal process of the palatine; Pf, pterygoid fossa.

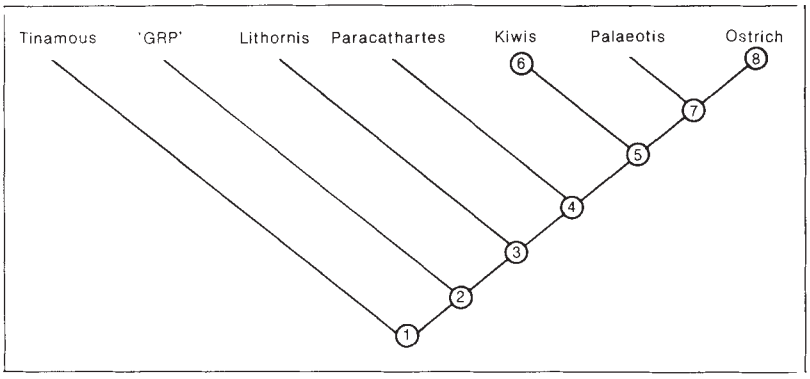


Fig. 2 Cladogram of some taxa of palaeognathous birds of unequal rank. Ratites without direct bearing to this discussion are omitted. Character suites are presumed to be apomorphous for all clades above 1. The caudal process of palatine is convergently absent in *Lithornis*, *Paracathartes* and Kiwis (not shown in cladogram). Character suites are: (1) palate palaeognathous, caudal process of palatine bone present, pterygoid narrow, vomer articulates with premaxillae, rhamphothecal grooves present, internarial septum small, splenial extends near to or forms part of mandibular symphysis, bill narrow, cervical vertebrae stout, ilioischiadic foramen open-ended, femur gracile, primary osteons of tibiotarsus predominantly oriented longitudinally, tibiotarsus with large lateral cnemial crest and well differentiated medial and lateral condyles, tarsometatarsus with pronounced, spherical intercotylar eminence, four toes present; (2) internarial septum large, pterygoid broad, quadrate fits into V-shaped notch in mandible, buttressed from behind by retroarticular process of mandible, scapula with styloid acromion, tarsometatarsus with large interosseous foramen, deeply concave anterior face and deeply excavated origin of extensor digiti IV brevis muscle; (3) pterygoid fossa present, vomer does not articulate with premaxillae; (4) primary osteons of tibiotarsus predominantly oriented transversely; (5) scapulocoracoid fused and sternum unkeeled, humerus long and spindly, with pronounced internal tuberosity and without pneumatic foramen, antebrachium short; (6) bill elongate, eyes reduced, tarsometatarsus without concave anterior face and excavated origin of extensor digiti IV brevis muscle, interosseous foramen small; (7) scapulocoracoid with large glenoid, acromion and acrocoracoid process, carpometacarpus with accessory foramina between each of the metacarpal bones, ilioischiadic foramen elongate, femur short, tarsometatarsus long, hallux absent (condition of quadrate mandible articulation unknown); and (8) pterygoid fossa absent, caudal process of palatine bone absent, V-shaped notch in mandible for quadrate absent, bill broad, cervical vertebrae elongate, femur short and robust, tibiotarsus with pronounced anterior cnemial crest and undifferentiated medial and lateral condyles, tarsometatarsus without spherical intercotylar eminence, concave anterior face and excavated origin of extensor digiti IV brevis muscle, interosseous foramen small and only two trochleae present.

gnathous tinamous (Tinamidae), the extant sister group of ratites[1,2,4,6]. *Palaeotis weigelti* was a crane-sized, flightless palaeognathous ostrich from the Middle Eocene of Germany, whose ancestry may be traced to the *Lithornis*-cohort. *Lithornis*[7], *Paracathartes*[8] and *Palaeotis*[9] were originally described from isolated, fragmentary remains and were mistaken for neognathous birds. Newly available, essentially complete skeletons that include the palatal bones (Fig. 1) now show unequivocally that these were palaeognathous birds. *Palaeotis* should not be confused with the incomplete fossils comprising the Asian family Ergilornithidae that have been suggested as ancestors of ostriches[10,11]. Ergilornithids were not palaeognathous birds.

The monophyletic origin of the *Lithornis*-cohort can be shown on the basis of shared derived characters (Fig. 2, character suite 2), all of which are absent in tinamous. Cranial (Fig. 1) and histological characters (Fig. 3) that characterize the living ratites evolved step by step in the phylogeny of the *Lithornis*-cohort. The large pterygoid fossa with no palatine process first appeared in clade 3, and the longitudinal pattern of bone osteons was replaced by the transverse pattern in clade 4 (Fig. 2). A higher-taxonomic group, comprised of the *Lithornis*-cohort, would thus be paraphyletic. That is, it does not include all of its evolutionary derivatives (the ratites), which have secondarily lost some of the distinguishing flight-related characters of the *Lithornis*-cohort.

An important element in the elucidation of ratite phylogeny is the fossil *Palaeotis*. Two skeletons of *Palaeotis weigelti* (specimen GM 4362, Geiseltalmuseum, Martin-Luther University, Halle a. Saale, GDR and specimen HLMD Me 7530, Hessisches Landesmuseum, Darmstadt, FRG) reveal all the characters that together define ratite birds,[1,2,6]. Derived characters that further distinguish *Palaeotis* as an ostrich (family Struthionidae) are

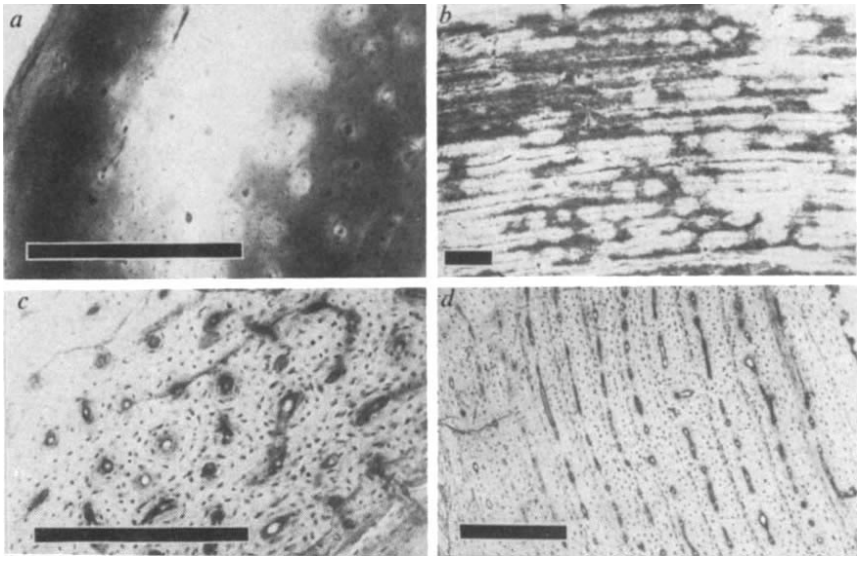


Fig. 3 Bright field photomicrographs of transverse lapidary thin sections through the same level (distal metaphysis) of *a*, tinamou (*Rhynchotus*, Princeton University PU 728); *b*, ratite (*Rhea*, specimen unnumbered); *c*, *Lithornis* (USNM 290554); and *d*, *Paracathartes* (USNM 361407), showing histological character states used in Fig. 2. Vascular canals appear as thin dark lines, spots, or circles on a light background of inorganic bone. Large dark regions are more opaque regions of bone, in which vascular canals may appear as light spots. Scale, 0.3 mm. The patterns formed by primary osteons are characteristically different in the leg bones of palaeognathous birds (tinamous and ratites) and neognathous birds (all other birds)[16,17]. The primary osteons in the tibiotarsus of *Lithornis* are predominantly parallel to the longitudinal axis of the bone, as in the tibiotarsus of the palaeognathous tinamous but unlike ratites and neognathous birds. Primary osteons in the tibiotarsus of *Paracathartes* are predominantly arranged in concentric rings that lie in the transverse plane of the bone, as in the tibiotarsus of ratites but unlike the tinamous and neognathous birds. Although the patterns of primary osteons could be evolved separately by unrelated birds, no correlation has been found between the pattern of primary osteons and flightlessness, ontogeny, or overall size among neognathous birds[18].

listed (Fig. 2, legend, character suite 7). Characters that are derived at the level of the *Lithornis*-cohort (Fig. 2, character suite 2) and primitively retained by *Palaeotis*, but not found in other birds, show that ostriches evolved from this group. Only *Palaeotis* and members of the *Lithornis*-cohort have palates with a caudal process of the palatine and a pterygoid fossa (Fig. 1), although both characters are not found together in *Lithornis*, *Paracathartes*, or the GRP'. *Palaeotis* forms an important link between the *Lithornis*-cohort and ostriches because all these characters are absent in large ratites, having been secondarily lost by all but kiwis.

Thus, ostriches were not only present in the Northern Hemisphere, but they may have evolved there long after the breakup of Gondwanaland. If so, then ostriches emigrated from Europe to Africa some time during the early or middle Tertiary, a dispersal route for which there are independent data for other taxa[12]. The hypothesis of ratite origins in Gondwanaland relies on the same route of dispersal, but in the opposite direction, to account for the presence of the recently extinct ostriches of Asia[2,13]. The absence of the *Lithornis*-cohort in early Tertiary deposits of the Southern Hemisphere does not conclusively prove that these birds lived only in the Northern Hemisphere at that time, but their absence there is conspicuous since *Lithornis* has been collected in relative abundance in the Northern Hemisphere for more than two centuries.

The vicariance biogeography hypothesis of ratite distribution requires that populations of flightless ratites were isolated as the result of geotectonic events and that the birds' own powers of dispersal were minimal. However, this hypothesis does not explain the occurrence of kiwis (Apterygidae) in New Zealand. Sibley and Ahlquist[4] could not reconcile the early (Cretaceous) divergence of the New Zealand plate from Gondwanaland with the relatively late (Eocene) divergence of kiwis from other ratite birds, based on DNA hybridization data. The fossils discussed here suggest that kiwis could have evolved directly from a palaeognath that flew to New Zealand and lost its power of flight independently of other ratite birds. Of the ratites, only kiwis retain primitive characters that are uniquely derived in the *Lithornis*-cohort (Fig. 1; Fig. 2, character suites 2 and 3), suggesting a relatively close relationship.

The far reaching implications of this study lie with the temporal calibration of the DNA hybridization molecular clock. Sibley and Ahlquist[4] postulated that the divergence of ostrich DNA from rhea DNA was initiated by, and therefore could be dated by, the spreading of the Atlantic seafloor. They have since[14] applied this calibration to a variety of other avian and mammalian taxa. With the simplicity of the vicariance biogeography hypothesis of ratite origins now challenged, one cannot accept this part of the calibration of the DNA molecular clock without at least some degree of scepticism. I have elsewhere[15] discussed problems with the four other data used to calibrate the DNA molecular clock.

I thank A. J. Charig for his comments on this manuscript and D. Baird, J. Becker, D. P. Domning, A. Elzanowski, P. D. Gingerich, H. Haubold, A. E. Houde, W. v. Koenigswald, S. L. Olson, A. de Ricqles, G. Stanton and R. L. Zusi for their assistance. This work was supported by the A. Wetmore Memorial Fund of the American Ornithologists Union, A. Wetmore Memorial Fund of the Smithsonian Institution, F. M. Chapman Memorial Fund of the American Museum of Natural History, Museum of Paleontology of the University of Michigan, National Science Foundation Dissertation Grant (BSR 8313209), Sigma Xi, Smithsonian Predoctoral Fellowship and National Institutes General Medical Sciences Grant (GM07803) to L. V. Leak of Howard University.

Received 23 January; accepted 18 September 1986.

1. Bock, W. J. *Proc. XIII Int. Congr. Orn.* 39-54 (1963).
2. Cracraft, J. *J. Zool.* **169**, 455-545 (1973).
3. Prager, E. M., Wilson, A. C., Osuga, D. T. & Feeney, R. E. *J. molec. Evol.* **8**, 283-294 (1976).
4. Sibley, C. G. & Ahlquist, J. E. *Evolution Today, Proc. Second int. Congr. Syst. evol. Biol.* (eds Scudder, G. G. E. & Reveal, J. L.) 301-335 (Carnegie-Mellon University, Pittsburgh).
5. Houde, P. & Olson, S. L. *Science* **214**, 1236-1237 (1981).
6. Parkes, K. C. & Clark, G. A. Jr *Condor* **68**, 459-471 (1966).
7. Owen, R. *Proc. Geol. Soc. Lond.* **3**, 162-166 (1840).
8. Harrison, C. J. O. *Tertiary Res. spec. Pap.* **5**, 7-10 (1979).
9. Lambrecht, K. *Jb. hallesch. Verb. Erforsch. Mitteldt. Bodenschatze* **7**, 1-11 (1928).
10. Feduccia, A. *The Age of Birds* (Harvard University Press, Cambridge, 1980).
11. Olson, S. L. in *Avian Biology* Vol. 8 (eds Farner, D. S., King, J. R. & Parkes, K. C.) 80-238 (Academic, New York, 1985).
12. Van Couvering, J. A. & Van Couvering, J. A. H. *Proc. VI Congr. Brataslava, Reg. Com. Mediterranean Neogene Strat.*, 363-367 (Brataslava, Czechoslovakia, 1975).
13. Cracraft, J. *Ibis* **116**, 494-521 (1974).
14. Sibley, C. G. & Ahlquist, J. E. *J. molec. Evol.* **20**, 2-15 (1984).
15. Houde, P. *Auk* (in the press).
16. Amprino, R. & Godina, G. *Anat. Anz.* **95**, 191-214 (1944).
17. Zavattari, E. & Cellini, I. *Monitore zool. ital.* **64**, 189-200 (1956).
18. Houde, P. *Auk* (in the press).

Phospholipid vesicles as a model system for biomineralization

Stephen Mann*, John P. Hannington* & R. J. P. Williams†

* School of Chemistry, University of Bath, Bath BA2 7AY, UK
† Inorganic Chemistry Laboratory, South Parks Road, Oxford OX1 3QR, UK

The tailoring of inorganic minerals such as iron oxides for functional use in biological systems for iron storage[1], structural support[2] and magnetoreception[3] involves biological regulation of crystal structure, particle size, morphology and crystallographic organization. The encapsulation of crystallochemical reactions within enclosed biological microvolumes enables control to be exerted over: (1) the chemical regulation, by passive or facilitated ion-transport, of localised supersaturation levels (2) the stereochemical requirements for ion-binding, redox and nucleation events at the organic matrix interface and (3) the spatial organization of crystal growth and morphology[4,5]. Matrix-mediated growth of inorganic materials has not been systematically investigated *in vitro* even though it may have applications in crystal engineering and materials science[6]. We have used phospholipid unilamellar vesicles of ~300 Å diameter to study membrane-mediated processes of iron oxide crystal growth. Intravesicular deposits differ in structure, morphology and size from precipitates formed from reactions in bulk aqueous solution. Mediating factors include vesicle shape and dimension, diffusion-limited processes of ion-transport and ion-binding at the curved lipid headgroup surface. These results help elucidate biomineralization and have technological relevance to the controlled synthesis of monodisperse sols with catalytic and magnetic properties.

Figure 1 illustrates the general reaction scheme adopted in our work. Unilamellar phospholipid vesicles with diameters of ~300 Å were prepared from phosphatidylcholine (17 mM) sus-

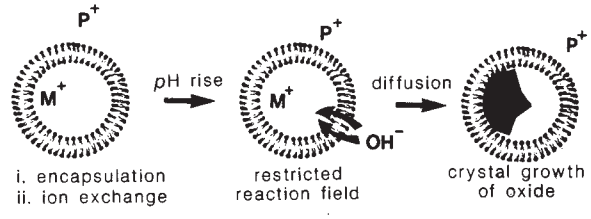


Fig. 1 Reaction scheme for membrane-mediated crystal growth of metal oxides in phospholipid vesicles. Cations (M^+) are encapsulated by sonication and replaced in the external phase by inert cations (P^+) by ion-exchange chromotography. This procedure results in organised restricted reaction fields for crystal growth. Increases in extravesicular *p*H result in OH^- influx and the nucleation and growth of metal oxides within the intravesicular space.

Table 1 Electron diffraction data for intravesicular precipitates

d spacing (Å) and associated Miller indices Intravesicular solution (see text)		
[Fe(III)]	[Fe(II)]	[Fe(II)/Fe(III)]
2.6 (021)	2.93 (220)	2.5 (110)
2.08 (140)	2.52 (311)	1.95 (113)
1.8 (211)	2.06 (400)	1.49 (106)
1.49 (250)	1.75 (422)	
1.32 (321)	1.44 (440)	
	1.24 (622)	
	1.05 (800)	
	1.00 (822)	
	0.92 (840)	

Lines of strong intensity are underlined. Miller indices are shown in parentheses.

pensions at 4 °C in the presence of Fe(II) and/or Fe(III) solutions. Three preparative routes were investigated: encapsulation in the presence of (1) 100 mM $FeCl_3$ at *p*H 2.0, (2) 100 mM $FeCl_2 4H_2O$ at *p*H 4.0, (3) 50 mM $FeCl_2$ and 50 mM $Fe(NO_3)_3$ solution at *p*H 2.0. In each experiment, metal ions external to the vesicles were removed by ion-exchange chromatography (Na-form). Precipitation of Fe(II) and/or Fe(III) ions in the restricted reaction field delineated by the vesicle membrane was then induced by addition of aqueous NaOH (final *p*H, 12.0) to the extravesicular solution. A pale yellow colouration was noted in all the solutions after 30 min and it increased in intensity over several days. No precipitation was observed after bench centrifugation. By contrast, in similar precipitation reactions in the absence of vesicles, rapid precipitation of iron oxides was apparent. Unstained samples were air dried onto electron microscope grids and studied by transmission electron microscopy, selected area electron diffraction and X-ray microanalysis.

Electron micrographs of iron-containing vesicles before *p*H increase and precipitation showed that the intravesicular iron was not present as a simple salt solution but that extensive adsorption via ion-binding at the choline headgroups had taken place (Fig. 2*a*). Precipitation of the intravesicular iron after NaOH addition resulted in the formation of discrete, finely divided particles with mean diameters 35–40 Å (Fig. 2*b*). X-ray microprobe analysis of vesicles before and after precipitation showed the presence of Fe, P and Cl (trace). None of these elements was detected in background analyses of the electron microscope grid. Table 1 shows the results obtained from electron diffraction identification of these intravesicular deposits. Intravesicular Fe(III) solutions crystallized as poorly-ordered spherulitic geothite (α-FeOOH) whereas extended aggregates of ferrihydrite were formed in bulk solution. Precipitation of intravesicular Fe(II) solutions eventually produced discrete 30–50 Å diameter magnetite (Fe_3O_4) crystallites (Fig. 2*b*) whereas bulk oxidation produced a mixture of acicular lepidocrocite (γ-FeOOH) and goethite. Oxidation and precipitation of initial intravesicular Fe(II)/Fe(III) solutions formed poorly-ordered ferrihydrite in contrast with the crystallization of 100–500 Å magnetite particles in the absence of vesicles.

These results can be rationalized in terms of the control by the vesicle membrane on the rate of OH^- diffusion into the intravesicular space. We have shown elsewhere[7] that the influx of hydroxyl ions into phospholipid vesicles is critically related to the efflux of counter-anions from the intravesicular volume. The net rate of OH^- diffusion and hence the rate of iron oxide formation will be dependent on the efflux rate of the slower-moving chloride anion[8]. The rate of intravesicular Fe(II) oxidation, which is dependent on the rate of increase of intravesicular *p*H, has been followed spectrophometrically by measurement of the Fe(III) absorption band at 420 nm. Whereas bulk oxidation at neutral or alkaline *p*H is rapid and follows overall

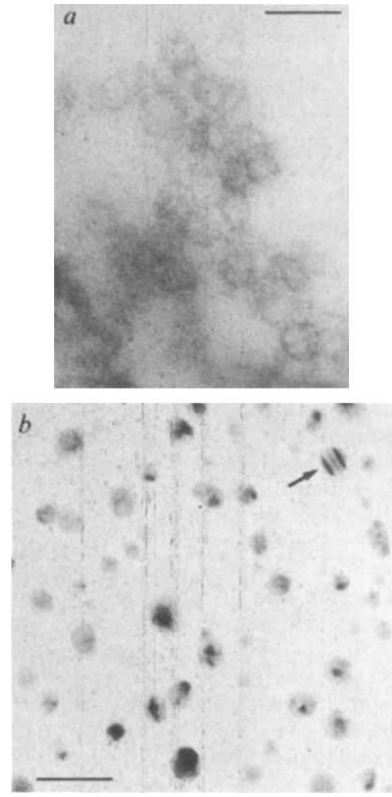


Fig. 2 *a*, Transmission electron microscope (TEM) image of phospholipid vesicles containing Fe(III) ions at *p*H 2.0. The binding of Fe(III) at the charged headgroups results in the *in situ* staining of the inner membrane surface. Bar, 60 nm. *b*, High magnification TEM image of intravesicular magnetite particles. Some particles show electron dense banding (arrow) which may arise from domains in twinned orientation or from diffraction contrast due to local crystal bending. Bar, 15 nm.

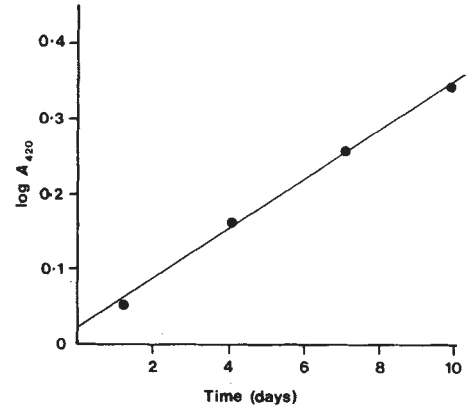


Fig. 3 Plot of log absorbance at 420 nm (A_{420}) against time for the intravesicular oxidation of Fe(II) solutions. The linear plot represents an overall first-order rate dependence. Samples were stored in the dark at 4 °C between readings. TEM studies showed minimal evidence of vesicle degradation over the course of the kinetic experiments.

third-order kinetics at constant $[O_2]$ (ref. 9), intravesicular Fe(II) oxidation shows overall first-order behaviour ($k = 3.7 \times 10^{-7}\ s^{-1}$) (Fig. 3) indicative of a diffusion-controlled mechanism.

These kinetic restrictions favour the formation of thermodynamically stable products such as geothite from Fe(III) sol-

utions, rather than ferrihydrite which is formed under conditions of rapid precipitation. In addition, the retardation of the rate of intravesicular Fe(II) oxidation due to the slow increase in intravesicular *p*H favours the formation of mixed valence oxides such as magnetite whereas rapid oxidation at high *p*H in bulk solution generates hydrous iron(III) oxides such as goethite and lepidocrocite. The formation of intravesicular ferrihydrite rather than the bulk precipitated magnetite from 1:1 Fe(II)/Fe(III) solutions reflects a similar mediation of crystal growth by the membrane. The slow increase in intravesicular *p*H will allow oxidation and favour the preferential precipitation of the Fe(III) component at acid *p*H values.

Finally, particle sizes of the above intravesicular products are determined by the vesicle internal diameter and the encapsulation efficiency. In the experiments, assuming a simple salt solution and 100% encapsulation efficiency, ~2,000 Fe irons would be present per vesicle. This would result, for example, in intravesicular magnetite particles of 25 Å diameter. Our particles are slightly larger than this, which can be accounted for by additional ions arising from saturation binding of Fe at the headgroup sites (~1,500 on the inner surface, Fig. 2*a*). We also note that all the intravesicular particles are spherical or disk-shaped whereas bulk precipitates such as γ- and α-FeOOH are acicular. Nucleation of the intravesicular solids, induced at the lipid interface binding sites, is presumably such that the initial stages of crystal growth are partially constrained by the curvature of the vesicle membrane.

In conclusion, we have demonstrated the mediation of crystal-growth processes of iron oxides in the presence of organized membrane-bound assemblies. Future work will extend this aspect of membrane-mimetic chemistry to investigate the role of headgroup spacing, charge density and curvature on the nucleation and crystal growth of inorganic solids which are biologically and industrially relevant.

Received 1 September; accepted 17 October 1986.

1. Ford, G. C. *et al. Phil Trans. R. Soc.* B**304**, 551-565 (1984).
2. Mann, S., Perry, C. C., Webb, J. & Williams, R. J. P. *Proc. R. Soc.* B**227**, 179-190 (1986).
3. Mann, S., Frankel, R. B. & Blakemore, R. P. *Nature* **310**, 405-407 (1984).
4. Mann, S. *Struct. Bond.* **54**, 127-174 (1983).
5. Mann, S. *J. inorg. Biochem.* (in the press).
6. Birchall, J. D. *Trans. Br. Ceram. Soc.* **83**, 158-165 (1984).
7. Mann, S., Kime, M. J., Ratcliffe, R. G. B. & Williams, R. J. P. *JCS Dalton Trans.* 771-774 (1983).
8. Mann, S. & Williams, R. J. P. *JCS Dalton. Trans.* 311-316 (1983).
9. Stumm, W. & Morgan, J. *Aquatic Chemistry* (Wiley, New York, 1984).

A thylakoid processing protease is required for complete maturation of the lumen protein plastocyanin

Johan Hageman*†, Colin Robinson‡, Sjef Smeekens*† & Peter Weisbeek*†

* Department of Molecular Cell Biology and
† Institute of Molecular Biology, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands
‡ Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

Plastocyanin, a photosynthetic electron carrier functional in the chloroplast lumen, is synthesized in the cytosol as a precursor (preplastocyanin) with an amino-terminal transit sequence[1]. This transit peptide contains the information specifying import into and routing within the chloroplasts and is removed in at least two steps[2]. An intermediate is observed in the stroma after the transport of preplastocyanin through the chloroplast envelope; mature plastocyanin is present in the lumen, after transport over the thylakoid membrane. We show here that the stromal processing protease[3] is not responsible for both processing events. It cleaves the precursor protein only to the intermediate size and a novel protease located in the thylakoids processes this intermediate to the mature protein. This second protease recognizes the processing intermediate but not the precursor. Thus plastocyanin import involves cleavage at the intermediate processing site mediated by the stromal protease and then cleavage at the mature processing site mediated by the thylakoid protease.

Most chloroplast proteins are synthesized in the cytosol and subsequently transported into the organelle by a post-translational, ATP-dependent mechanism (for reviews see refs 4 and 5). The proteins are initially synthesized as larger precursors containing amino-terminal pre-sequences ('transit sequences') that are removed after import into the organelle. Transit sequences are believed to contain some, if not all, of the information specifying transport into the chloroplast and localization within the subchloroplast compartment[2,6,7]. Chloroplasts are composed of three membranes enclosing three discrete soluble phases—the intermembrane space, the stroma and the lumen. The proteins that are imported into the chloroplast have to cross one or more of these membranes; plastocyanin, a thylakoid lumen protein[8] must cross all three *en route* to its site of function. The import of preplastocyanin takes place in two stages[2], indicating that its transit sequence can be divided into an amino-terminal chloroplast-import domain and a thylakoid-transfer domain, each with its own processing site. We tried to determine whether the

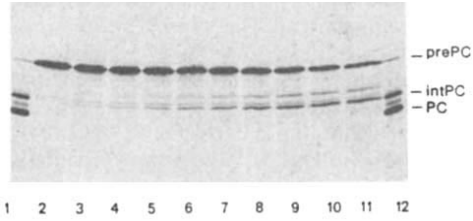


Fig. 1 Processing of plastocyanin precursor by stromal processing protease. Lane 1, import marker; lanes 2-11, incubation of plastocyanin precursor with partially purified protease for 0, 5, 10, 15, 20, 30, 40, 60, 90 and 120 minutes; lane 12, import marker.
Methods. Plastocyanin was synthesized in a wheat-germ system with RNA generated by *in vitro* transcription of the *Eco*RI-linearised pSPPC74 plasmid in an SP6 system[2]. The transcripts were capped by the addition of methylated cap ($m^7G(5')ppp(5')G$). The RNA was translated in a wheat-germ translation system in the presence of ^{35}S-methionine (1,000 Ci $mmol^{-1}$)[14]. The stromal protease was concentrated ~350-fold from stromal extracts of pea chloroplasts as described[3]. The processing incubation at 27 °C contained 10 μl translation mix, 30 μl processing buffer (100 mM HEPES-KOH *p*H 8.5, 220 mM KCl, 5 mM $MgCl_2$) and 40 μl processing enzyme. At the times indicated, 8 μl aliquots were removed and put into a mixture of 32 μl H_2O and 40 μl sample buffer (125 mM Tris-HCl *p*H 7.6, 4% SDS, 5% sucrose, 5% 2-mercaptoethanol) and boiled for 2 min. The products were analysed by SDS-15% polyacrylamide gel electrophoresis using the Laemmli buffer system[15] and fluorography. Import marker: chloroplasts were isolated from young pea plants (cv Feltham First) as described[16] and finally taken up in import buffer (50 mM HEPES-KOH *p*H 8.0, 330 mM sorbitol) at a chlorophyll concentration of 1 mg ml^{-1}. In the import experiment 400 μl chloroplasts were incubated with 100 μl precursor plastocyanin translation mix, 25 μl 5× import buffer and 80 μl 200 mM methionine at 25 °C in the light[16]. After 60 min the chloroplasts were diluted with 10 ml import buffer and reisolated by centrifugation. The pellet was taken up in 600 μl sample buffer without protease treatment and boiled for 2 min. For each import marker lane, 15 μl of this mixture was used. The minor band between the processing intermediate and the mature plastocyanin may be an artefact as result of the length of the incubation.

```
        -66          -50      -40      -29      -20      -10      -1
prePC   MATVTSSAAVAIPSFAGLKASSTTRAATVKVAVATPRMSIKASLKDVGVVVAATAAAGILAGNAMA
PC58                                          MSIKASLKDVGVVVAATAAAGILAGNAMA
```

Fig. 2 Amino-acid sequence of the transit peptides of the wild-type plastocyanin precursor (prePC)[1] and the substitute intermediate (PC58). The amino acids are numbered relative to the mature processing site. The intermediate processing site is around the isoleucine residue at −27.

stromal processing protease is responsible for both the processing of the precursor to the intermediate and the intermediate to the mature form, or if not, what other proteases are involved.

Preplastocyanin was made by transcription *in vitro* of the *Silene* plastocyanin complementary DNA[2] in an SP6 vector[9], followed by translation of the RNA in a wheat-germ translation system. This procedure yields a single labelled band of precursor (Fig. 1 lane 2). The two minor lower relative molecular mass bands immunoprecipitate with plastocyanin antibodies (data not shown) and probably represent in-frame internal translation initiations at the methionine (ATG) codon at position −29 and at the isoleucine (ATC) codon at position −27 of the preplastocyanin transit sequence (see also ref. 2). The frequencies of these internal initiations vary between different wheat-germ systems (data not shown). The stromal processing enzyme was isolated from pea chloroplasts and purified ~350-fold[3]. As homologous size markers, mature plastocyanin and the processing intermediate were obtained from an *in vitro* import experiment with preplastocyanin (see Fig. 1 legend). The import marker lanes used in Figs 1, 2 and 4 show mature plastocyanin, a small amount of non-imported precursor and two intermediate bands the upper one of which represents the stromal processing intermediate[2].

Incubation of preplastocyanin synthesized *in vitro* with the partially purified stromal protease shows that the stromal processing protease is able to convert the precursor into a molecule the same size as the stromal intermediate (Fig. 1). It cannot be excluded that this processing occurs in two or more steps as the band between the precursor and the final processing product is consistently observed. The stromal protease however is not able to accomplish complete processing: even after prolonged incubation no mature protein is detected. (See also Fig. 4, lane 2.)

These results with *Silene* precursor plastocyanin differ from previous results obtained with barley and wheat plastocyanin[3]. We re-examined the processing of wheat plastocyanin precursor, using immunoprecipitated wheat plastocyanin mature protein as a reference during electrophoresis, instead of purified pea plastocyanin. This revealed that the protein that resulted from the processing with the stromal protease actually migrated more slowly than the immunoprecipitated mature plastocyanin from wheat (data not shown). The stromal protease therefore also processes the wheat plastocyanin precursor only to the intermediate size.

These results also imply that a second protease is responsible for the processing of the intermediate to the mature protein. Because the most likely location for this enzyme was the thylakoids, we incubated precursor protein with a thylakoid preparation. To release the putative protease from the thylakoids the membranes were disrupted by treatment with detergent[10]. Because of the possibility that precursor plastocyanin in addition to the precursor would not be recognized, a substitute intermediate protein was used to probe for the second protease. For the *in vitro* synthesis of this substitute intermediate a plastocyanin clone (PC58) was used which has had the initiator ATG codon removed by exonuclease digestion. Because of this, translation of *in vitro* transcribed *pc*58 RNA in a wheat-germ system starts at the first ATG, the methionine codon at position −29 downstream in the transit sequence (Fig. 2). Consequently PC58 protein is only a few amino acids longer than the stromal

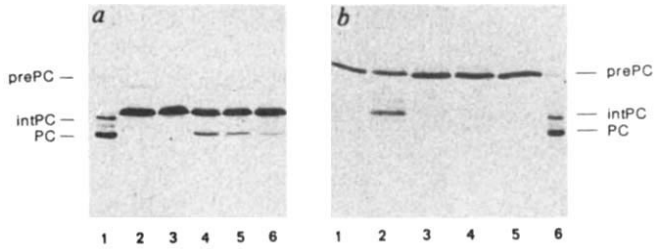


Fig. 3 A second protease is released from the thylakoids after Triton X-100 treatment. Incubation of PC58 (*a*) and plastocyanin precursor (*b*) with intact and Triton X-100-treated thylakoids. Panel *a*: lane 1, import marker; lane 2, PC58 translation mix; lane 3, incubation of PC58 with intact thylakoids; lanes 4–6, incubation of PC58 with thylakoids that were treated with 0.2, 0.5 and 1.0% Triton X-100 respectively. Panel *b*: lane 1, precursor plastocyanin translation mix; lane 2, incubation of preplastocyanin with intact thylakoids; lanes 3–5, incubation of preplastocyanin with thylakoids that were treated with 0.2, 0.5 and 1.0% Triton X-100 respectively; lane 6, import marker.

Methods. *In vitro* synthesis of PC58 was performed as described for plastocyanin precursor in Fig. 1 legend. The plasmid used is called pSPPC58. Thylakoids were isolated from isolated pea chloroplasts; intact chloroplasts at a chlorophyll concentration of 1 mg ml^{-1} in import buffer were pelleted by centrifugation and lysed by the addition of 20 mM Tris-HCl, *p*H 8.0, to the original volume. Aliquots of lysed chloroplasts were transferred to Eppendorf tubes and centrifuged in an Eppendorf centrifuge for 5 min at 4 °C. The pellet which contained the thylakoids was resuspended in the original volume of 20 mM Tris-HCl, *p*H 8.0. For the Triton X-100 treatment the thylakoids were pelleted again and taken up in solutions of 20 mM Tris-HCl, *p*H 8.0, containing the desired percentages of Triton X-100. The incubations (at 25 °C) contained 1 μl translation mix, 10 μl 20 mM Tris-HCl, *p*H 8.0, 10 μl processing buffer (see Fig. 1) and 70 μl untreated or Triton X-100-treated thylakoids. Reactions were stopped by adding 100 μl SDS-sample buffer followed by boiling for 2 min.

intermediate. PC58 is not imported into chloroplasts *in vitro* (data not shown). *In vitro* synthesis of PC58 yields one major labelled protein (Fig. 3*a* lane 2). The minor band above the PC58 protein probably represents an in-frame initiation at the isoleucine codon (position −55). The relative efficiency of initiation at this codon also depends on the wheat-germ extract used.

Formation of mature plastocyanin was not observed after incubation of PC58 protein or plastocyanin precursor protein with intact thylakoids (Figs 3*a*, lane 3; Fig. 3*b*, lane 2). PC58 was not processed at all, although the precursor was processed to the intermediate size. The latter processing was identical to that observed after incubation with stromal protease (Fig. 1) suggesting that the thylakoid fractions were not free from stromal contamination. Consistent with this, the SDS-polyacrylamide gels of the thylakoid preparation showed bands characteristic of ribulose bisphosphate carboxylase, the major stromal protein (not shown). PC58 is not processed by the stromal protease (data not shown) and therefore is not affected by stromal contamination.

Disruption of the thylakoid membranes by mild Triton X-100 treatment releases a protease activity that converts PC58 into a protein the same size as mature plastocyanin (Fig. 3*a* lanes 4–6). This processing activity is specific for the intermediate, as preplastocyanin is not processed at all under these conditions (Fig. 3*b* lanes 3–5), showing that the stromal protease contamination of the thylakoid fraction is completely inhibited by Triton X-100. The thylakoid protease is only moderately sensitive to this detergent. It can also be relased from the thylakoids by sonication (data not shown).

Finally, the combination of the stromal and the thylakoid protease is sufficient for the complete processing of preplastocyanin. After the incubation of the precursor first with the

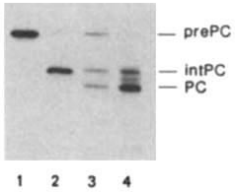


Fig. 4 Processing of plastocyanin precursor to the mature size by successive processing with a stromal protease and thylakoid protease. Lane 1, plastocyanin precursor translation mix; lane 2, incubation with stromal protease for 2 h; lane 3, incubation with stromal protease for 1 h followed by Triton X-100-treated (0.2%) thylakoids for 1 h; lane 4, import marker.
Methods. Plastocyanin precursor synthesized *in vitro* was incubated with partially purified stromal protease (see Fig. 1) for 1 h. Half was then incubated for 1 h with an equal volume of thylakoid vesicles treated with 0.2% Triton X-100 (see Fig. 3). The remainder was incubated for 1 h, after which samples were analysed by SDS-polyacrylamide gel electrophoresis followed by fluorography.

partially purified stromal protease and then with Triton X-100-treated thylakoids, a protein the same size as mature plastocyanin is detected (Fig. 4, lane 3). Prolonged incubation with stromal protease only resulted in a further decrease of the amount of precursor and an increase of the amount of intermediate (Fig. 4, lane 2), confirming that the stromal protease is not able to complete the processing. Processing of wheat preplastocyanin with this combination of proteases also results in complete processing to the mature size (data not shown).

The involvement of two distinct proteases in the processing of plastocyanin precursor to the mature size emphasizes the special nature of the transport route taken by nuclear-encoded proteins that are functional in the lumen. It also confirms the division of the plastocyanin transit sequence into a chloroplast-import domain and a thylakoid-transfer domain[2]. Partial processing of preplastocyanin by the stromal protease is consistent with the observation from the *in vitro* reconstitution experiments that the processing intermediate is found in the stroma, whereas the mature protein is found in the lumen[2]. The specificity of the thylakoid protease for the intermediate form suggests that this protease recognizes the plastocyanin transit sequence only after the removal of the chloroplast-import domain by the stromal protease. The determination of the exact location of the thylakoid protease is under investigation, and the first results on the purification show that the protease is water-soluble and retains activity on separation from the thylakoid membrane.

Plastocyanin, or more specifically the plastocyanin intermediate, is not the only protein that is imported from the stroma into the thylakoid lumen. A mechanism has to exist for lumen proteins made in the chloroplast too. By analogy with the proposed evolutionary origin of the chloroplast[11], this probably involves a prokaryote-like import mechanism. It will be interesting to determine whether plastocyanin and proteins synthesized in the chloroplast share import mechanisms or not, especially considering that the thylakoid-transfer domain remarkably resembles a prokaryote signal peptide[12]. The thylakoid protease would then function analogously to the signal peptidase[13].

This work is supported in part by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO) and by an EMBO short term fellowship to J.H.

Received 1 June; accepted 17 October 1986.

1. Smeekens, S., De Groot, M., Van Binsbergen, J. & Weisbeek, P. *Nature* **317**, 456-458 (1985).
2. Smeekens, S., Bauerle, C., Hageman, J., Keegstra, K. & Weisbeek, P. *Cell* **46**, 365-375.
3. Robinson, C. R. & Ellis, R. J. *Eur. J. Biochem.* **142**, 337-342 (1984).
4. Ellis, R. J. *A. Rev. Pl. Physiol.* **32**, 111-137 (1981).
5. Cashmore, A. *Bio/Technology* **3**, 803-808 (1985).
6. Van den Broeck, G. *et al. Nature* **313**, 358-363 (1985).
7. Schreier, P. H., Seftor, E. A., Schell, J. & Bohnert, H. J. *EMBO J.* **4**, 25-32 (1985).
8. Boulter, D., Haslett, B. G., Peacock, D., Ramshaw, J. A. & Scawen, M. D. in *Plant Biochemistry II* (ed. Northcote, D.) 1-40 (University Press, Baltimore, 1977).
9. Melton, D. A. *et al. Nucleic Acids Res.* **12**, 7035-7056 (1984).
10. Siefermann-Harms, D., Ross, J. W., Kaneshiro, K. H. & Yamamoto, H. Y. *FEBS Lett.* **149**, 191-196 (1982).
11. Margulis, L. *Origin of Eukaryotic Cells* (Yale University Press, 1970).
12. Von Heijne, G. *EMBO J.* **3**, 2315-2318 (1984).
13. Benson, S. A., Hall, M. N. & Silhavy, T. J. *A. Rev. Biochem.* **54**, 101-134.
14. Anderson, C. W., Straus, J. W. & Dudock, B. S. *Meth. Enzym.* **101**, 635-644 (1983).
15. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
16. Grossman, A. R., Bartlett, S. G., Schmidt, G. W., Mullett, J. E. & Chua, N.-H. *J. biol. Chem.* **257**, 1558-1563 (1982).

Cholinergic neurones acquire adrenergic neurotransmitters when transplanted into an embryo

James Niles Coulombe & Marianne Bronner-Fraser

Developmental Biology Center and Department of Developmental and Cell Biology, University of California, Irvine, California 92717, USA

During development, cells become progressively restricted, until they reach their final phenotype. Differentiation was originally thought to be irreversible, but phenotypic plasticity has been observed in a variety of cell types, for example sympathetic neurones[1], the limb blastema[2] and some glial cell types[3]. A detailed description of the individual steps that lead to expression or reversal of phenotype is essential to understand the molecular events underlying cell differentiation. We examined whether ciliary neurones acquire adrenergic properties when exposed to a permissive embryonic environment. Cholinergic neurones were selectively labelled with a retrogradely transported marker and injected into chick embryos during active neural crest migration[4]. Four to five days after injection, some of the labelled neurones were found in 'adrenergic sites' and had developed catecholamine histofluorescence. The cells had thus accumulated adrenergic neurotransmitters even after differentiation into cholinergic neurones. This result shows that neurotransmitter plasticity occurs in cholinergic neurones and suggests that the neurotransmitter phenotype can be modified by the embryonic environment.

In the autonomic nervous system, both adrenergic and cholinergic neurones are derived from the neural crest[5]. In the differentiation of neural crest-derived neurones, the choice of neurotransmitter is a critical decision for proper autonomic function since the neurotransmitters used by sympathetic and parasympathetic neurones have antagonistic effects on their target organs. There is evidence that some neural crest cells may be fated to become cholinergic neurones; early migrating cranial neural crest cells have choline acetyltransferase activity[6] and acetylcholinesterase[7]. In addition, a monoclonal antibody to ciliary neurones recognizes a subpopulation of the neural crest cells in the mesencephalon[8]. After differentiation, the local environment may selectively promote survival of those neurones which contain the appropriate neurotransmitter, and/or instruct the developing cells to differentiate in the appropriate direction. Experiments showing that neurotransmitter choice in sympathetic neurones is labile even in post-mitotic cells[1] support the latter idea. Superior cervical ganglion cells (which normally produce catecholamines) produce acetylcholine[1,9-11] and form cholinergic synapses[12,13] when grown in the presence of a diffusible factor made by nonneuronal cells. Evidence for neurotransmitter plasticity in the autonomic nervous system is also found *in vivo* in the cholinergic sympathetic innervation to the sweat gland of the rat footpad[14-16] and in the cholinergic innervation of the iris following sympathectomy[17].

We examined neurotransmitter plasticity in cholinergic neurones, to ascertain whether cholinergic neurones are 'fated'

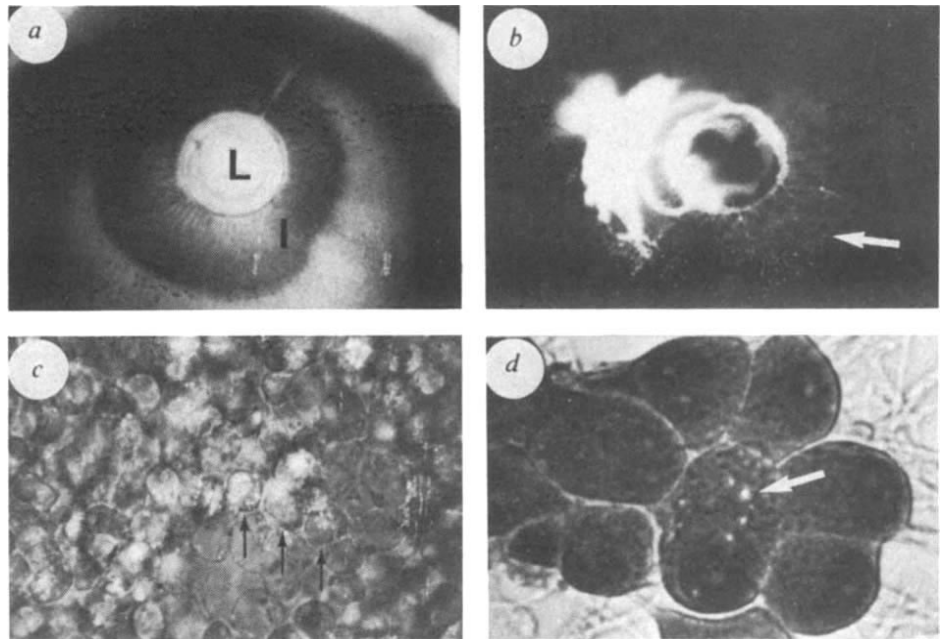


Fig. 1 Photomicrographs showing: *a*, *b*, the site of injection for retrogradely transported fluorescent latex microspheres in the iris region (I) of the eyeball. *c*, microsphere-labelled neurones in the ganglion. *d*, retrogradely labelled ciliary neurone. *a*, Transmitted light view of the iris region of a 6.5 day quail eye; L, lens (15×). *b*, Epifluorescence view of the same eye as *a*. Arrow, site where fluorescent latex microspheres filled the capillary bed of the ciliary body where axons from ciliary ganglion neurones terminate (15×). c, Combined fluorescence and bright field micrograph of a squashed ciliary ganglion showing fluorescent latex microspheres within neuronal cell bodies (596×). *d*, Bright field superimposed with fluorescence photomicrograph of a latex-based-labelled ciliary neurone (indicated by arrow) that has ChAT immunoreactivity (140×). All such neurones were found to be ChAT positive.

Methods. For labelling, the eyes were removed, leaving the ciliary ganglion in place at the back of the eyeball. The venous sinus (the canal of Schlemm), which supplies the ciliary ganglion and iris region with blood was severed ~3 mm proximal to the iris to avoid artefactual transport. Fluorescent latex microspheres (0.5–2.0 μm spheres, latex impregnated with rhodamine)[20] were pressure injected through the canal of Schlemm into the capillary bed that supplies the iris and ciliary body, the target sites for ciliary ganglion neurones. Because the venous sinus was severed, the neurones from the ciliary ganglion were only in contact with the latex microspheres at their growth cones and synapses, from which the microspheres were endocytosed and transported to the cell body. The eyeballs were incubated 8–10 h in a 95% oxygen, 5% CO_2 atmosphere at 37 °C in culture medium consisting of Eagle's minimum essential medium (Gibco) containing 15% horse serum, 10% chick embryo extract and 0.02% gentamycin sulphate. For ChAT immunocytochemistry, the microsphere-labelled ganglia were fixed in paraformaldehyde and picric acid, infiltrated with sucrose, and cryostat sectioned. Sections were dried onto gelatin coated microscope slides, and reacted overnight with primary antibodies against choline acetyltransferase (provided by Dr Miles Epstein) or tyrosine hydroxylase (Eugene Tech International Inc.). Antibody binding was made visible by treatment with biotin conjugated second antibodies, a biotin-avidin-horseradish peroxidase complex (Vectastain ABC), development with diamino-benzidine, and intensification with osmium. For cell counts, cell nuclei were stained with 4-6-diamidino-2-phenylindole dihydrochloride (DAPI).

to express only cholinergic neurotransmitters or labile with respect to neurotransmitter content. Previous investigators have transplanted cholinergic ganglia into presumptive adrenergic regions of the chick embryo[18]. In these transplants, the non-neuronal cells of the ganglia survived and synthesized catecholamines in the host adrenergic sites; the neurones, however, did not survive[19]. To test the transmitter plasticity of cholinergic neurones *in vivo* adequately, an experimental paradigm requires: (1) a transplantation method which supports neuronal survival; and (2) an unequivocal marker for the implanted cholinergic neurones. Here, we describe a new approach that meets these requirements.

To obtain a selectively labelled population of cholinergic neurones, we removed the eyes from 6.5 day quail embryos, leaving the ciliary ganglion in place at the back of the eyeball. Fluorescent latex microspheres were injected into a venous sinus supplying the iris and ciliary body, target sites for the neurones from the ciliary ganglion (Fig. 1). Microspheres were endocytosed by neuronal synapses and growth cones from ciliary neurones with processes at the target site, and were then retrogradely transported to the soma. The fluorescent latex microspheres, which were easily identified by their intense, punctate yellow-red fluorescence, have been shown to remain in the neuronal cell bodies for long periods of time and not to damage the neurone[20]. In 6.5 day old quail ciliary ganglia, ~15% of all cells were labelled with fluorescent microspheres. The labelled population represented ~20% of the neurones in the ganglia. The major advantage of this labelling technique is that only differentiated neurones, with processes at the target site, were labelled by retrograde transport.

To confirm that only neurones were labelled by the fluorescent microspheres, ciliary ganglia were dissociated and the cells were probed with an antiserum that recognized neurofilament proteins (Figure 2 and ref. 21). We classified a cell as labelled if it contained at least 5 fluorescent microspheres in its cytoplasm. Microsphere labelling was only found in cells containing neurofilament-like immunoreactivity. In addition, the microsphere-labelled cells had large cell bodies characteristic of neurones and many also had axonal processes.

An important assumption in our experiments is that the labelled neurones are cholinergic before transplantation. Four day old quail ciliary gangia have measurable choline acetyltransferase (ChAT) activity[22] and ciliary ganglion neurones *in vitro* form cholinergic synapses[23]. Recently, however, some tyrosine hydroxylase (TH) immunoreactive cells have been detected in avian ciliary ganglia[24]. Labelled ciliary ganglia were immunostained with antibodies against ChAT or TH to determine which of the neurotransmitter biosynthetic enzymes was present in labelled neurones. ChAT immunoreactivity (Fig. 1*d*) was found in all latex-bead-labelled neurones identified ($n > 1000$). In fact, all cells with the large cell bodies characteristic of neurones appeared to have ChAT immunoreactivity. Less than 1% of ciliary ganglion cells (estimated at 0.5% in chicken ciliary ganglia[25], and 0.8% in quail (unpublished data)) have TH immunoreactivity. No latex-bead-labelled neurones were detected among this population of TH-immunoreactive cells ($n = 104$).

Before implantation, the labelled ganglia were dissociated using conditions that promote neuronal survival in tissue culture[23]. The viability of the labelled neurones after dissociation was confirmed by their ability to regenerate severed axons and

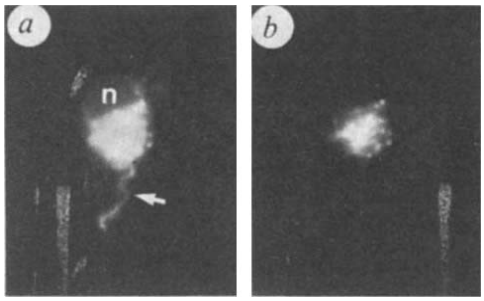


Fig. 2 Fluorescent photomicrographs of latex-bead-labelled ciliary ganglion neurones with anti-neurofilament antibodies. *a*, A ciliary ganglion neurone viewed with epifluorescence through a fluorescein filter has set neurofilament-like immunoreactivity. This cell has the large cell body and axonal process (arrow) characteristic of neurones. Due to fluorescence crossover, some latex microspheres could be seen through this filter set. Note the cell nucleus (n) 144×). b, The same neurone as in *a* viewed with a rhodamine filter shows numerous fluorescent latex microspheres in the cytoplasm (1440×). Diffuse signal is due to microspheres which were out of the plane of focus. Antibodies to neurofilament proteins were found to stain all cells containing fluorescent microspheres, suggesting that the microsphere-labelled cells were, in fact, neurones.

Methods. After labelling, ciliary ganglia were removed from the eyeball and cleaned of connective tissue and the ganglion capsule. The ganglia were then dissociated in 0.1% trypsin (Difco) in Modified Puck's Saline with glucose (MPG). Following trypsinization, the ganglia were mechanically dissociated by gentle trituration. For some experiments, the dissociated cells were filtered through a Nitex 15 μm screen. The dissociated cell suspension was fixed in 4% paraformaldehyde in 0.1 M phosphate buffer for 30 min at 22 °C. After washing, the cells were permeabilized in MPG containing 0.05% saponin with 5% goat serum for 30 min and reacted at 4 °C overnight with a rabbit primary antiserum (diluted 1:200) that specifically recognizes neurofilament proteins[21]. The cells were then washed and reacted with fluorescein-conjugated goat antibodies against rabbit immunoglobulin G (IgG) (diluted 1:20) for 1-2 hours at 37 °C.

to survive in culture. For some experiments, the ganglion cell population was enriched for neuronal cells by preplating[26]. The suspension containing labelled cholinergic neurones was implanted into the trunk somites of stage 12-18 chicken embryos using an injection technique[4], which reproducibly placed the cells on the ventral neural crest migratory pathway[27]. Embryos were incubated for an additional 4-5 days after injection, during which time some endogenous neural crest cells differentiated into catecholamine-containing cells[28]. The hosts were then rapidly frozen, processed for catecholamine-histofluorescence, and serially sectioned. During the 4-5 day incubation period, labelled neurones moved along the ventral neural crest pathway and collected around the dorsal aorta, where endogenous neural crest cells coalesce to form the adrenal medulla, sympathetic ganglia and aortic plexuses. Thus the labelled, dissociated ciliary ganglion neurones survived after implantation and were primarily located in or near sites normally populated by neural crest cells.

Injected embryos were examined for cells containing both fluorescent microspheres and catecholamines. We have identified, in nine separate embryos, a total of 33 microsphere-labelled ciliary ganglion neurones which also contained catecholamine histofluorescence (Figure 3*b*,*c*). Using a Feulgen stain, we have confirmed in several cases that the microsphere-labelled catecholamine-containing neurones also had the quail heterochromatin marker[29]. The number of labelled neurones containing adrenergic neurotransmitters represents ~5-10% of the microsphere-labelled neurones identified in section and ~50% of the labelled neurones that were incorporated into adrenergic structures. The other microsphere-labelled neurones were generally found near adrenergic cells but were not incorporated into adrenergic structures. Before injection no ciliary ganglion neurones contained catecholamines[25] or had a catecholamine uptake system[30]. Thus, some cholinergic neurones gained adrenergic neurotransmitters in the proper embryonic environment.

We observed a larger percentage of labelled neurones from ciliary ganglia in the adrenomedullary cords than in sympathetic ganglia. A similar distribution pattern to that of the labelled ciliary ganglion neurones is observed when inert latex beads are injected into the embryos[31]. Both the neurones and the beads may be moved to ventral sites by an environmentally controlled mechanism independent of cell motility. Only a fraction of the labelled neurones had catecholamine histofluorescence, and all of these were located in adrenergic structures such as the sympathetic ganglia or adrenomedullary cords. Localization in adrenergic structures may therefore be a prerequisite for the adrenergic phenotype. This suggests that the local environment, rather than the migratory process, influences the expressed phenotype of the neurones. A possible explanation for the small number of labelled neurones that were incorporated into host adrenergic structures may be that they are 'outcompeted' from these sites by endogenous neural crest cells. Because we have only examined the catecholamine content of the labelled ciliary ganglion neurones, our experiments cannot discriminate between uptake or synthesis of catecholamines. Neurones from ciliary ganglia *in vitro* do not have a catecholamine uptake system (refs 25, 30 and unpublished observations). Therefore, the labelled neurones have at least expressed biochemical systems for uptake and storage of catecholamines.

Control transplantations were performed to determine whether microsphere label could be passed to previously unlabelled cells after transplantation. Labelled ganglion cells were killed by repeated freeze/thawing, yielding microsphere-labelled ganglion-cell debris; the debris was then coinjected with an equal volume of a suspension of unlabelled live ciliary ganglion cells as in the experiments described above. In these controls, no cells containing multiple microspheres were identified. Individual microspheres were sometimes observed either in cells or extracellularly, but multiple microspheres were not acquired by unlabelled cells. This validates our selection of 5 microspheres as the criterion for labelling.

Thus, under appropriate conditions, cholinergic neurones can acquire an important characteristic of the adrenergic phenotype. Other investigators have observed that the environment encountered by neural crest cells can influence their subsequent differentiation. In quail/chick chimaeras, premigratory neural crest cells from presumptive cholinergic regions of the neural axis become adrenergic when grafted into the trunk[32]. In tissue culture, altering the environmental conditions can affect the differentiation of neural crest cells[33,34]. The culture environment can also induce explanted ciliary ganglion neurones to express some adrenergic characteristics; ciliary neurones *in vitro* can express immunocytochemically detectable tyrosine hydroxylase[24,25]. Similar results were observed *in situ* where some fibres from adult ciliary neurones apparently acquire tyrosine hydroxylase immunoreactivity after sympathectomy[17]. But, no catecholamines were detected in the ciliary neurones either *in vivo* after sympathectomy or *in vitro*, which contrasts with our findings.

Our results, together with the previous observations of the adrenergic to cholinergic transition[1,9-16], suggest that autonomic neurotransmitter plasticity is bidirectional. It remains to be determined whether the accumulation of catecholamines is due to uptake or synthesis, whether the transition can occur only in a defined period, whether cells reenter the cell cycle before expressing catecholamines and whether the labelled neurones can express both cholinergic and adrenergic traits simultaneously. Because transplanted cholinergic neurones can

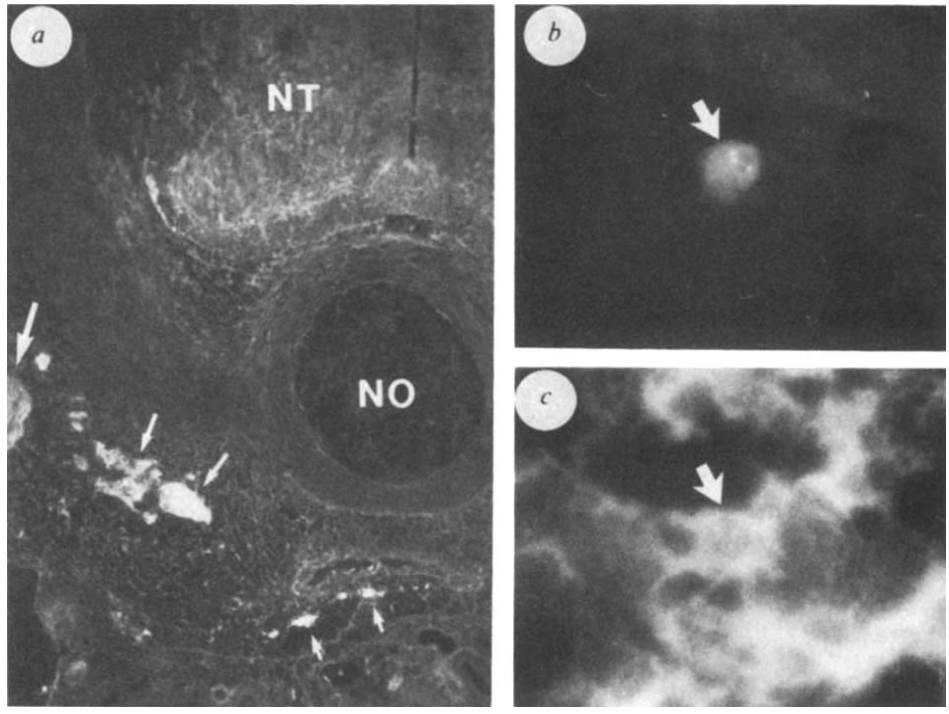


Fig. 3 Transverse sections through 7.5-day-old chicken embryos processed for catecholamine histofluorescence 5 days after labelled ciliary ganglion neurones were injected into the trunk somites. *a*, Low magnification of a transverse section through the trunk illustrating the sites where neural crest cells have formed catecholamine-containing derivatives. NT, neural tube; NO, notochord; arrows, sympathetic ganglia, adrenomedullary cells, and aortic plexuses (81×). *b*, A high power photomicrograph of the same embryo showing a retrogradely labelled neurone located in the adrenomedullary cords, viewed through a filter set designed for rhodamine fluorescence. Fluorescent latex microspheres, which have an intense punctate yellow-red fluorescence, can be visualized in the cell indicated by the arrow (890×). *c*, The same field as in *b* viewed through a catecholamine filter set, through which adrenergic cells are identified by their intense blue-green fluorescence. The microsphere-labelled cell (indicated by arrow) contains catecholamines and was part of a group of adrenergic cells located in the adrenomedullary cords (890×).

Methods. Embryos were rapidly frozen in liquid nitrogen-cooled isopentane, freeze-dried at −40 °C for ≥4 days, equilibrated to room temperature and exposed to formaldehyde vapour for 1-2 h[35]. Fixed embryos were embedded under vacuum in paraplast, serially sectioned at 10 μm and placed dry onto albumin or acrylic adhesive coated microscope slides. Sections were covered with microscope immersion oil (Nikon) and examined at 800× under epifluorescent illumination. We confirmed that the same cell profile contained both microspheres and catecholamine histofluorescence using both rhodamine and catecholamine filter sets and examining the cell in a number of focal planes.

accumulate catecholamines, our results are consistent with a cell lineage model in which the neurotransmitter phenotype of autonomic neurones is strongly influenced by instructive cues from the local environment.

We thank Dr Scott Fraser for his helpful comments on the manuscript and Georgia Guillory and Jeffrey D. Pratt for technical assistance. This work was supported by USPHS Grant HD-15527-01 and by Basic Research Grant 1-896 from the March of Dimes.

Received 27 June; accepted 18 September 1986.

1. Patterson, P. H. & Chun, L. L. Y. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3607-3610 (1974).
2. Brockes, J. P. *Science* **225**, 1280-1287 (1984).
3. Raff, M. C., Miller, R. H. & Noble, M. *Nature* **303**, 390-396 (1983).
4. Bronner, M. & Cohen, A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1843-1847 (1979).
5. Le Douarin, N. M. *The Neural Crest* (Cambridge University Press, London, 1982).
6. Smith, J., Fauquet, M., Ziller, C. & Le Douarin, N. M. *Nature* **282**, 853-855 (1979).
7. Cochard, P. & Colty, P. *Devl Biol.* **98**, 221-238 (1983).
8. Barald, K. F. *Neuronal Development* (ed. Spitzer, N. C.) 110-119 (Plenum, New York and London, 1982).
9. Patterson, P. H. & Chun, L. L. Y. *Devl Biol.* **57**, 268-280 (1977).
10. Patterson, P. H. & Chun, L. L. Y. *Devl Biol.* **60**, 473-481 (1977).
11. Potter, D. D., Landis, S. C., Matsumoto, S. G. & Furschpan, E. J. *J. Neurosci.* **6**, 1080-1098 (1986).
12. Johnson, M., Ross, D. *et al. Nature* **262**, 308-310 (1976).
13. Furschpan, E., MacLeish, P., O'Lague, P. H. & Potter, D. *Proc. natn. Acad. Sci. U.S.A.* **73**, 4225-4229.
14. Landis, S. & Keefe, D. *Devl Biol.* **98**, 349-372 (1983).
15. Yodlowski, M., Fredieu, J. R. & Landis, S. C. *J. Neurosci* **4**, 1535-1548 (1984).
16. Leblanc, G. & Landis, S. C. *J. Neurosci.* **6**, 260-265 (1986).
17. Bjorklund, H., Hokfelt, M., Goldstein, L., Terenus, L. & Olson, L. *J. Neurosci.* **5**, 1633-1640 (1985).
18. Le Douarin, N. M., Teillet, M. A., Ziller, C. & Smith, J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2030-2034 (1978).
19. Dupin, E. *Devl Biol.* **105**, 288-299 (1984).
20. Katz, L. C., Burkhalter, A. & Dreyer, W. J. *Nature* **310**, 498-500 (1984).
21. Drager, U. C., Edwards, D. L. & Kleinschmidt, J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6408-6412 (1983).
22. Smith, J., Fauquet, M., Ziller, C. & Le Douarin, N. M. *Nature* **282**, 853-855 (1979).
23. Nishi, R. & Berg, D. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5171-5175 (1977).
24. Teitelman, G. *et al. J. Neurosci.* **5**, 29-39 (1985).
25. Iacovitti, I. *et al. Devl Biol.* **110**, 402-412 (1985).
26. Collins, F. & Dawson, A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2091-2094 (1983).
27. Bronner-Fraser, M. *Devl Biol.* **113**, 44-55 (1986).
28. Enemar, A., Falck, B. & Hakanson, R. *Devl Biol.* **11**, 268-283 (1965).
29. Le Douarin, N. M. *Expl. Cell Res.* **77**, 459-468 (1973).
30. Rohrer, H. & Sommer, I. *J. Neurosci.* **3**, 1683-1693 (1983).
31. Bronner-Fraser, M. E. *Devl Biol.* **91**, 50-63 (1982).
32. Le Douarin, N. M. & Teillet, M. A. *Devl Biol.* **41**, 162-184 (1974).
33. Howard, M. & Bronner-Fraser, M. *J. Neurosci.* **5**, 3302-3309 (1985).
34. Perris, R. & Löfberg, J. *Devl Biol.* **113**, 327-341 (1986).
35. Falck, B. & Owman, C. *Acta Univ. lung.* **2**, 4-25 (1965).

Variable and conserved neutralization antigens of human immunodeficiency virus

Robin A. Weiss*, Paul R. Clapham*, Jonathan N. Weber*, Angus G. Dalgleish*, Lawrence A. Lasky† & Phillip W. Berman†

* Institute of Cancer Research, Chester Beatty Laboratories, Fulham Road, London SW3 6JB, UK
† Department of Molecular Biology, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080, USA

Human immunodeficiency virus type 1 (HIV-1, HTLV-III/LAV), the retrovirus responsible for acquired immune deficiency syndrome (AIDS), shows a high degree of genetic polymorphism, particularly in the *env* gene[1-6]. We have examined sera from rabbits and guinea pigs immunized with gp130, a recombinant *env* glycoprotein[7], and sera from HIV-1-infected subjects[8], to test their capacity to neutralize a panel of genetically divergent HIV-1 isolates. The sera raised against recombinant antigen specifically neutralized the virus strain from which the *env* gene was cloned (HTLV-IIIB), but not an independent isolate (HTLV-IIIRF). One rabbit serum tested on seven isolates cross-neutralized two at lower titres. In contrast, human sera from Britain and Uganda, chosen for ability to neutralize HTLV-IIIRF, cross-neutralized six other

HIV-1 isolates. When serum and isolate were derived from the same subject, the serum was in some cases effective at slightly lower concentrations (higher titres). Human complement did not affect neutralization titres. These findings indicate that genetically diverse HIV-1 isolates carry both variable and widely conserved antigenic epitopes for neutralizing antibodies. The identification of shared epitopes may help the development of protective vaccines.

The production of a glycosylated recombinant envelope antigen of HIV-1 (gp130) by hamster CHO cells was recently described[7]. This recombinant antigen has at its amino-terminus 25 amino acids of the herpes simplex type I glycoprotein D, followed by amino acids 61–531 of the *env* gene of the HIV-1 isolate HTLV-IIIB (IIIB) representing most of the major envelope glycoprotein, gp120. This antigen is secreted by CHO cells. Rabbits and guinea-pigs immunized with partially purified gp130 in Freund's adjuvant neutralize the infectivity of the IIIB isolate and immunoprecipitate not only recombinant gp130 but also gp160 and gp120 of IIIB (ref. 7). In Table 1 we confirm the neutralization of the IIIB isolate by two different assays, but find no significant neutralization of a distinct, Haitian isolate, HTLV-IIIRF (RF).

To ascertain whether antisera raised against recombinant gp130 specifically neutralized only the HIV-1 strain from which the recombinant was derived, envelope pseudotypes of vesicular stomatitis virus (VSV) were prepared from seven different HIV-I isolates, and their sensitivity to neutralization was tested with rabbit anti-gp130 serum. In addition, twelve sera from British and Ugandan[9] subjects pre-selected for neutralizing activity to the RF isolate were screened, as was an enzyme-linked immunosorbent assay (ELISA)-positive Ugandan serum that was inactive for RF neutralization. Pre-immune rabbit serum, anti-HIV-1 ELISA-negative human sera and a VSV(HTLV-1) pseudotype were employed as negative controls.

Table 2 shows that the rabbit anti-gp130 serum specifically neutralized IIIB, cross-neutralized MN (ref. 1) and MA2 (ref. 10) isolates, but did not neutralize four other isolates. In contrast, the human anti-HIV-1 sera, whether British or Ugandan, cross-neutralized nearly all HIV-1 isolates, including two sera (British 23, 24) taken from subjects who were infected after receiving a single HIV-positive blood transfusion and who had no other

Table 1 Strain-specific neutralization of HIV-I by sera raised against recombinant *env* polypeptide gp130

	Neutralizing titre			
	HIV-1		VSV(HIV-1)	
Serum	IIIB	RF	IIIB	RF
Rabbit				
25	50	—	50	—
26	10	—	10	—
27	50	—	50	—
Guinea-pig				
1	10	—	10	—
2	50	—	50	—
3	50	—	50	—

Two methods were used to titrate the neutralizing activity of rabbit and guinea-pig sera to HIV-1 gp130, after heat inactivation at 56 °C for 30 min. The first assay measured surviving tissue-culture infectious doses (TCID) of HIV-1, when serum dilutions were mixed with 10^5 TCID in culture medium (RPMI-1640 with 10% fetal calf serum) for 1 h at 43 °C and dilutions of the reaction mixture were then added to 10^5 C8166 assay cells. The cultures were maintained for 7 days and the presence of large multinucleated syncytia was scored on day 4 and 7. HIV-1 end-point titrations ($TCID_{50}$) scored by appearance of syncytia in C8166 cells were equivalent to those scored by immunofluorescence of HIV-1 antigens in the cells or the presence reverse transcriptase in the medium of C8166 or H9 cells infected with HIV-1 (unpublished data). The neutralizing titre shown is the reciprocal of the highest dilution inactivating ≥90% TCID of HIV-1. The second assay[8] used the neutralization of vesicular stomatitis virus (VSV) pseudotypes bearing the envelope glycoproteins of HIV-1. Serum dilutions were mixed with 200 plaque forming units (p.f.u.) of VSV(HIV-1) for 1 h at 43 °C before plating in 30 mm diameter wells on 2.5×10^5 adherent CCRF-CEM cells pretreated with DEAE-dextran. After adsorption for 1 h, 10^6 CCL64 mink cells were added, followed by an agar overlay 2 h later. VSV cytopathic plaques were counted 20–48 h after VSV(HIV-1) plating, and the neutralizing titre is expressed as the reciprocal of the highest serum dilution giving ≥80% reduction of p.f.u. Pre-immune sera from the same animals had no neutralizing activity. The (−) sign indicates no significant neutralization with 1:9 serum dilution.

Table 2 Cross-neutralization of HIV-1 isolates by recombinant *env* anti-serum, and British and Ugandan human sera

				Neutralizing titres to VSV(HIV-1)					VSV
Serum	ELISA	RF*	IIIB†	MN†	ARV-2†	MA2†	RUT‡	Z129§	(HTLV-1)
Rabbit									
pre-immune	—	—	—	—	—	—	—	—	—
anti-gp130	+	—	50	10	—	10	—	—	—
Human									
British 1	—	—	—	—	—	—	—	—	—
6	+	10	10	250	6,250	50	250	50	—
12	+	50	10	250	1,250	50	250	50	—
14	+	10	—	250	250	10	250	50	—
17	+	50	10	250	1,250	50	50	250	—
21	+	50	10	10	1,250	10	50	250	—
22	+	250	—	10	1,250	10	250	250	—
23	+	50	10	10	1,250	10	50	50	—
24	+	250	10	10	6,250	50	250	250	—
Uganda 2	—	—	—	—	—	—	—	—	—
17	+	50	10	50	1,250	50	10	250	—
29	+	—	—	—	—	—	—	—	—
40	+	250	10	—	1,250	10	250	250	—
42	+	50	10	10	6,250	250	250	50	—
48	+	50	10	10	1,250	250	250	50	—

VSV(HIV-1) pseudotypes were prepared as described[8] for all 7 HIV-1 isolates. They were propagated in H9 cells except for Z129 in the K12 clone of HUT78 cells. Serial dilutions of heat-inactivated sera were reacted with 200 p.f.u. of VSV(HIV-1) and the surviving p.f.u. titrated as described for Table 1. VSV(HTLV-1) was grown in HOS/PL cells as described[13], and was used as a control to demonstrate the specificity of VSV(HIV-1) neutralizing activity.

Countries of origin: * Haiti; † USA; ‡ Tanzania; § Zaire.

Table 3 Titration of serum neutralizing activity for autologous and heterologous HIV-1 isolates

Serum	ELISA	Neutralizing titre to VSV(HIV-1)						
		RF*	ARV-2†	ARV-4†	RUT‡	Z84§	Z129§	IIIB†
B1	—	—	—	—	—	—	—	—
RF	+	250	250	NT	250	50	NT	10
ARV-2	+	50	1,250	NT	50	NT	NT	10
ARV-4	+	50	1,250	250	250	10	NT	50
RUT	+	—	10	—	50	—	10	—
Z84	+	10	250	NT	50	—	NT	10
Z129	+	10	50	NT	50	NT	—	10

NT, not tested.
Countries of origin: * Haiti; † USA; ‡ Tanzania; § Zaire.

known exposure to HIV-1. The Ugandan serum that was negative for RF neutralization failed to neutralize any isolate. It appears that the sera of humans infected with HIV-1 are more cross-reactive for neutralization than the rabbit anti-serum raised against gp130. However, as the neutralizing titre of the rabbit anti-serum was only 1/50 for the homologous virus (IIIB), a low degree of cross-reactivity might not have been detected. Higher titre anti-gp130 sera would merit investigation.

Indirect immunofluorescence techniques were used on H9 cells productively infected with IIIB, RF and ARV-2 to test whether the rabbit anti-gp130 serum recognized non-neutralizing epitopes of HIV-1. Positive results were obtained with each isolate with anti-gp130. It is clear, therefore, that some common epitopes on the virion envelope identifiable by membrane immunofluorescence are not targets for neutralization.

Table 2 also shows that VSV pseudotypes prepared with certain HIV-1 isolates (such as ARV-2) are generally more sensitive to neutralizing sera than other isolates (such as IIIB). This property does not appear to correlate with absolute VSV(HIV-1) titre, as all pseudotype preparations contained $\sim 10^4$ plaque-forming units per ml.

We previously reported[8] that some sera had high titres for weak neutralization (50% plaque reduction) but that sera taken from the same patients as the isolates (LAV-1, CBL-1) did not show preferential neutralization of their own isolate. However, sequence analysis has indicated that the *env* sequences of the four isolates tested show <3% nucleotide divergence. We therefore tested the neutralizing activity of sera against HIV-1 isolates from the same patients, using HIV-1 isolates that are known to be genetically polymorphic (Table 3). The data indicate that some virus isolates are more sensitive to neutralization by their autologous serum (for example, RF) and that some weakly neutralizing sera (for example, RUT) neutralize the autologous isolate best. But Z84 and Z129 were not neutralized by autologous sera.

The strong membrane immunofluorescence seen with some sera which show weak or no neutralizing activity against HIV-1 (ref. 8) suggests that non-neutralizing antibodies bind to virion envelopes. We therefore investigated whether complement enhanced the neutralizing activity of antisera directed to HIV-1 envelope glycoproteins. Our results indicate that neutralizing titres are not significantly enhanced by human complement (not shown).

Our experiments reveal the presence of both common and strain-specific epitopes in the target antigens for antisera responsible for the neutralization of HIV-1. The studies with recombinant gp130 (ref. 7) and VSV(HIV-1) pseudotypes indicate that the major viral envelope glycoprotein, gp120, is a target for neutralizing antibodies, but that does not exclude gp41 and possible *gag* gene products. Where the *env* sequences are known (IIIB, ARV-2 and RF), they show up to 15% nucleotide sequence divergence[3,4], and it is likely that the Zairian and Tanzanian isolates are even more divergent[5]. HIV-1, like the lentiviruses of ungulates[11], shows considerable genomic polymorphism, particularly in those regions of the *env* gene encoding domains not characteristic of oncoviruses[6].

Although some of the genetic variability exhibited by HIV-1 is reflected in divergence of neutralization antigens, there are insufficient data as yet to determine to what extent the variants have been immunologically selected. The human sera may be more cross-reactive than those from animals immunized with non-infectious antigen because of exposure to different epitopes, (1) through processing of newly made viral antigens, (2) opportunities for infection by multiple strains, and (3) the possible generation of variants during the course of infection[12]. Two serum samples were from subjects who had been exposed to HIV-1 on only one occasion (blood transfusion), but that does not exclude the possibility of infection with more than one strain present in the donor's blood. It is most unlikely that all the British and Ugandan sera would cross-neutralize American, Haitian, European and African strains through exposure to multiple, variable epitopes of viral antigen. We therefore conclude that certain neutralization antigens of HIV-1 are conserved between the diverse strains.

The variation of HIV-1-neutralizing antigens is in striking contrast to HTLV-1, for which neutralizing sera and virus isolates from Japanese, British West Indian and United States subjects are indistinguishable, though distinct from HTLV-2 (refs 13, 14). However, the neutralizing serotypes of avian leukosis viruses show considerable variation even within a single envelope subgroup[15,16], and different isolates of the bovine leukosis virus share some neutralizing epitopes of the gp51 envelope glycoprotein[17].

Our demonstration of conserved neutralization antigens, possibly those close to the binding site of gp120 to the CD4 cellular receptor[18-20] or other functionally conserved domains, may be important in the development of vaccines that elicit humoral and cellular immunity to HIV-1 infection. The strain specificity of the sera raised against recombinant gp130 reported here and those raised against purified gp120 of HTLV-IIIB (ref. 21) may indicate immunodominant determinants specific to the IIIB isolate. Further studies using monoclonal antibodies, and sera raised against different *env* constructs are in progress. By using a panel of VSV pseudotypes enveloped by glycoproteins of genetically and immunologically variant isolates of HIV-1, antisera raised against recombinant glycoproteins or synthetic peptides can be rapidly screened to identify those eliciting antibodies cross-reactive for neutralization.

We thank R. C. Gallo and M. Popovic for providing HIV-1 isolates, HTLV-IIIB, RF, MN and serum RF, and HTLV-1 transformed C91/PL and C8166 cells; L. Thiry and J. A. Levy for ARV-2 and ARV-4 viruses and sera; and D. Zagury for Z84 and Z129 viruses and sera. Human sera from Uganda were provided by R. G. Downing and N. K. Sewankambo. D. Whitby provided technical assistance. This study was supported jointly by the Cancer Research Campaign and the MRC. J.N.W. is supported by a clinical research fellowship from the Wellcome Trust.

Received 25 July; accepted 13 October 1986.

1. Wong-Staal, F. *et al. Science* **229**, 759-762 (1985).
2. Benn, S. *et al. Science* **230**, 949-951 (1985).
3. Hahn, B. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 4813-4817 (1985).
4. Starcich, B. R. *et al. Cell* **45**, 637-648 (1986).
5. Alizon, M., Wain-Hobson, S., Montagnier, L. & Sonigo, P. *Cell* **46**, 63-74 (1986).
6. Coffin, J. M. *Cell* **46**, 1-4 (1986).
7. Lasky, L. A. *et al. Science* **233**, 209-212 (1986).
8. Weiss, R. A. *et al. Nature* **316**, 69-72 (1985).
9. Serwadda, D. *et al. Lancet* **ii**, 849-852 (1985).
10. Webster, A. D. B. *et al. Lancet* **i**, 581-583 (1985).
11. Haase, A. *Nature* **322**, 130-136 (1986).
12. Hahn, B. H. *et al. Science* **232**, 1548-1553 (1986).
13. Clapham, P. R., Nagy, K. & Weiss, R. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2886-2889 (1984).
14. Hoshino, H. *et al. Int. J. Cancer* **36**, 671-675 (1985).
15. Ishizaki, R. & Vogt, P. K. *Virology* **30**, 375-387 (1966).
16. Weiss, R. A. & Biggs, P. M. *J. natn. Cancer Inst.* **49**, 1713-1725 (1972).
17. Bruck, C., Portetelle, D., Mammerickx, M., Mathot, S. & Burny, A. *Leukemia Res.* **8**, 315-321 (1984).
18. Dalgleish, A. G. *et al. Nature* **312**, 763-767 (1984).
19. Klatzmann, D. *et al. Nature* **312**, 767-769 (1984).
20. McDougal, J. S. *et al. Science* **231**, 382-385 (1986).
21. Matthews, T. *et al. Proc. natn. Acad. Sci.* U.S.A. (in the press).

Cytotoxic T lymphocytes recognize influenza haemagglutinin that lacks a signal sequence

A. R. M. Townsend*, J. Bastin*, K. Gould† & G. G. Brownlee†

* Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK
† Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

A surprising feature of most cytotoxic T lymphocytes (CTL) responding to influenza infection is that they recognize the unglycosylated (non-transmembrane) proteins of the virus, including the nucleoprotein[1–4]. Recognition of cells that express nucleoprotein by CTL does not depend on a definite signal sequence within the protein[5], and the epitopes recognized can be defined with short synthetic peptides *in vitro*[6]. Haemagglutinin (HA), the major transmembrane protein of the virus, is recognized by a minor population of CTL from infected mice. We have deleted the sequence coding for the N-terminal signal peptide from a complementary DNA encoding HA of the H1 subtype. The signal-deleted HA is detected with antibodies as a short-lived, unglycosylated, intracellular protein. However, CTL raised to the complete molecule recognize cells expressing the signal-deleted HA and vice versa. These results cast doubts on the assumption that CTL recognize the HA molecule only after its insertion into the plasma membrane.

Intuitively, proteins that are naturally inserted into the plasma membrane should not require a special mechanism for recognition by CTL. The requirement for a hydrophobic signal peptide at the N-terminus of the HA molecule for its translocation into the lumen of the endoplasmic reticulum, insertion into the membrane, glycosylation and transport to the cell surface[7,8] formed the basis for our experiment. If CTL recognize haemagglutinin only after it has been inserted into the plasma membrane then cells expressing signal-deleted HA should not be recognized by HA-specific CTL. If an independent mechanism exists for presentation of HA in an appropriate form at the cell surface, then deletion of the signal sequence should not affect recognition by CTL.

The DNA encoding the HA signal sequence was deleted by oligonucleotide-mediated mutagenesis from a full-length DNA copy of RNA segment 4 from influenza A/PR/8/34 (H1 subtype, Mount Sinai strain, see ref. 2) as described[9] (Fig. 1*a*). We chose to express the signal-deleted HA from the 7.5K early promoter in recombinant vaccinia[10]. Such HA-vaccinia can prime mice efficiently for a haemagglutinin-specific CTL response[11,12], and express high levels of HA protein in cell lines *in vitro*. The recombinant viruses here are referred to as L⁻H1-VAC (expressing the signal-deleted HA of the H1 subtype), H3-VAC (expressing the full-length HA of the H3 subtype from A/NT/60/68 virus) and H1-VAC (expressing the full length H1 haemagglutinin from the same promoter).

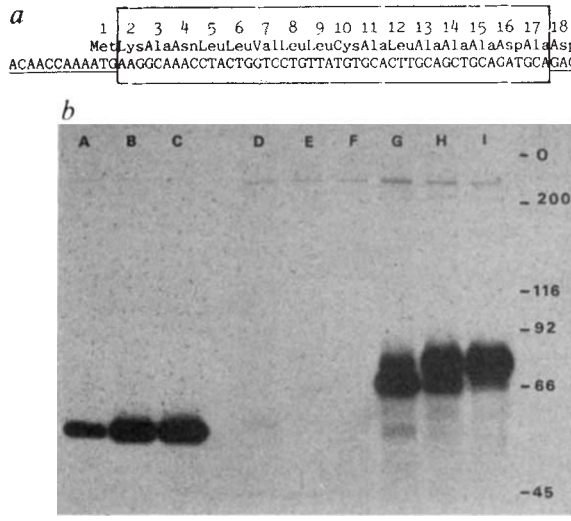


Fig. 1 *a*, Deletion of the H1 haemagglutinin (HA) signal sequence. *b*, Immunoprecipitation of signal deleted haemagglutinin. *a*, The N-terminal sequence of the H1 HA is from Winter *et al.*[13]. Sequence of the mutagenic oligonucleotide (24 bases) is underlined and the deleted segment coding for amino acids 2–17 is boxed. Mutagenesis was performed as described[9]. The mutant cDNA was sequenced by primer extension. The signal deleted H1 mutant and the full length H3 cDNAs were cloned into the vaccinia expression plasmid pGS62. Insertion into the vaccinia *tk* gene, isolation and purification of recombinant viruses followed standard procedures[10]. *b*, Murine L cells at 5×10^6 per ml were infected with 5–10 plaque forming units (PFU) of recombinant vaccinia virus per cell in RPMI containing 10% fetal calf serum (FCS) (RPMI/10) for 1.5 h at 37 °C. The cells were then washed twice in phosphate-buffered saline (PBS) and resuspended at 10^6 per ml in RPMI/5 for 3 h, washed twice in PBS and resuspended at 5×10^6 per ml in methionine-free RPMI, 5% FCS for 1 h before adding 120 μCi of ^{35}S-methionine for 30 min. They were then diluted to 10^6 cells per ml in RPMI/10 containing 4 mmol l^{-1} unlabelled L-methionine and the first aliquot of cells removed for lysis (time 0). The remainder were incubated at 37 °C and subsequent aliquots removed after 1 and 4 h. Immunoprecipitation was performed as described[5] on aliquots of 2×10^6 cells infected with L H1-VAC and NP-VAC (expressing the A/PR/8/34 nucleoprotein, used here as a specificity control), and 7×10^5 cells infected with H1-VAC. Reduced immunoprecipitates were electrophoresed on a 10% SDS polyacrylamide gel which was then treated with Amplify (Amersham), dried and exposed to pre-flashed X-ray film for 8 days. Lanes A–C, NP-VAC-infected cells immunoprecipitated with monoclonal antibody 469/6 (directed against nucleoprotein) at 0.1 and 4 h respectively after pulse labelling with ^{35}S-methionine. Lanes D–F, L⁻H1-VAC-infected cells immunoprecipitated with H36-18-2 (anti-H1 haemagglutinin), time points as for A–C. Lanes G–I, H1-VAC-infected cells immunoprecipitated with H36-18-2, time points as for A–C.

The mutant gene, lacking the codons for amino acids 2–17 inclusive, should code for an unglycosylated protein of relative molecular mass 61.6K[13]. Figure 1*b* shows the results of a pulse-chase immunoprecipitation experiment using monoclonal antibody H36-18-2 that binds both the signal-deleted and full-length haemagglutinin. Immediately after a 30 minute pulse with ^{35}S-methionine, the mutant HA is weakly detectable in infected L cells at the expected position for unglycosylated HA (~60K, lane D). The full-length molecule migrates at about 68K, consistent with rapid core glycosylation (lane G). During the four hour chase period the mutant HA becomes undetectable, imply-

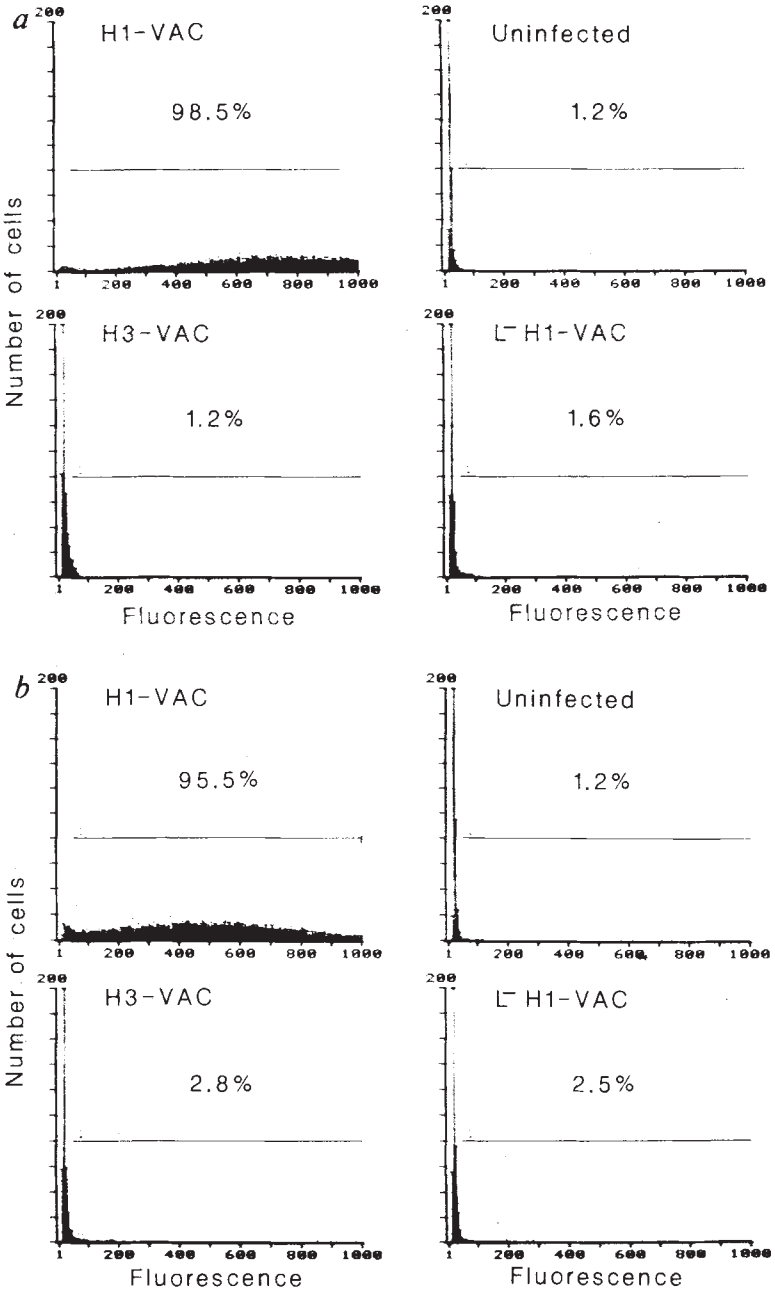


Fig. 2 Indirect immunofluorescence staining of L⁻H1-VAC- and H1-VAC-infected L cells. L cells were infected as described in Fig. 1 with either L⁻H1-VAC, H1-VAC, H3-VAC (included as a specificity control) or left uninfected. After 1.5 h cells were washed twice and resuspended at 10^6 per ml in RPMI/10 for 3 h at 37 °C. Aliquots of 5×10^5 cells were then stained by indirect immunofluorescence as described[5] *a*, The first layer was H36-18-2 monoclonal antibody to H1 haemagglutin at 5 $\mu g\ ml^{-1}$ after purification on protein A-Sepharose. The second layer was FITC-labelled affinity-purified goat anti-mouse antibody (Sigma) at 1:40 dilution. *b*, The first layer was hyperimmune rabbit anti-serum R401 raised to A/PR/8/34 virus diluted 1:2,000. The second layer was FITC-labelled affinity-purified goat anti-rabbit anti-serum (Sigma) diluted 1:100. The samples were analysed on an Ortho Cytofluorograf.

ing that it may be rapidly degraded (lanes D, E and F). In contrast the bulk of the full-length molecule progresses to a higher molecular mass of ~75K (lanes G, H and I). Control immunoprecipitations on cells infected with recombinant vaccinia expressing nucleoprotein (lanes A, B and C) gave a prominant band migrating at ~57K, as expected.

Figure 2 shows that the signal-deleted HA was not detectable at the cell surface with specific monoclonal (Fig. 2*a*) or polyclonal (Fig. 2*b*) antibodies. In contrast the full-length molecule was expressed at the surface of 98% of H1-VAC infected cells. In these experiments the target cells were infected for four and a half hours with recombinant vaccinia virus before being stained by indirect immunofluorescence. This is the equivalent of the mid time point in the ^{51}Cr release assay used to assess CTL recognition.

Recombinant vaccinia expressing the mutant and wild type HAs were then compared for their ability to prime CBA mice for a specific CTL response to the H1 subtype of influenza haemagglutinin. Animals primed *in vivo* with the L⁻H1-VAC were able to respond to wild-type HA expressed by A/PR/8/34 influenza virus *in vitro* as efficiently as mice primed with vaccinia expressing the full length HA. The response was shown to be specific because target cells expressing H1 but not H3 haemagglutinin were recognized (Fig. 3). Animals primed with H3-VAC (data not shown) did not make a detectable H1 haemagglutinin specific CTL response.

Finally, CTL from animals primed *in vivo* with A/PR/8/34 influenza virus were tested for their ability to recognize target cells expressing the signal-deleted HA. Most of the CTL in these polyclonal cultures are specific for conserved internal viral components that are shared by X31 (H3) and A/PR/8/34 (H1) influenza viruses. The subpopulation capable of recognizing haemagglutinin can be detected selectively with target cells expressing HA in the absence of other influenza proteins, after transfection or infection with a recombinant vaccinia[1,2,4,11,12].

Figure 4 (left panel) shows that the HA-specific CTL recognized target cells expressing the signal-deleted HA as efficiently as cells expressing the full-length molecule, despite the fact that

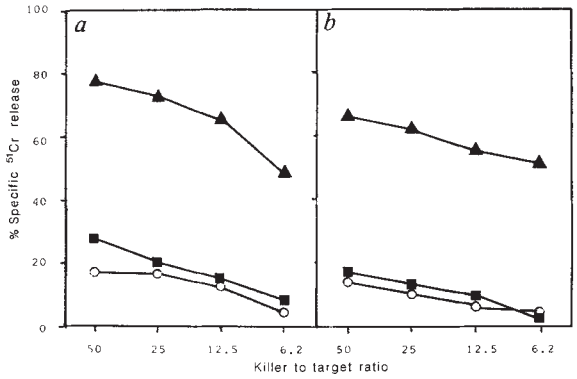


Fig. 3 The CTL response to infection by L⁻H1-VAC (*a*) and H1-VAC (*b*). Female CBA mice (obtained from NIMR Mill Hill), 3-6 months old, were primed intravenously by injecting 1×10^7 PFU of either L⁻H1-VAC or H1-VAC in 0.1 ml PBS into the tail vein as described[11]. After 12-14 days they were killed and their spleen cells stimulated *in vitro* with A/PR/8/34 influenza virus. After 5 days the responding cells were collected and tested by ^{51}Cr release assay on influenza-virus-infected L cells as described[1,2]. Responding CTL were incubated at the killer to target ratios indicated with ^{51}Cr-labelled target cells. Per cent specific ^{51}Cr release was calculated by (release by CTL − release in medium)/(2.5% Triton X-100 release − release in medium) × 100. Spontaneous release in medium from labelled L cells was 8-16% that released by Triton X-100. Target L cells were infected with; ▲, A/PR/8/34 (H1 subtype haemagglutinin); ■, A/NT/60/68 (H3 subtype haemagglutinin); ○, uninfected.

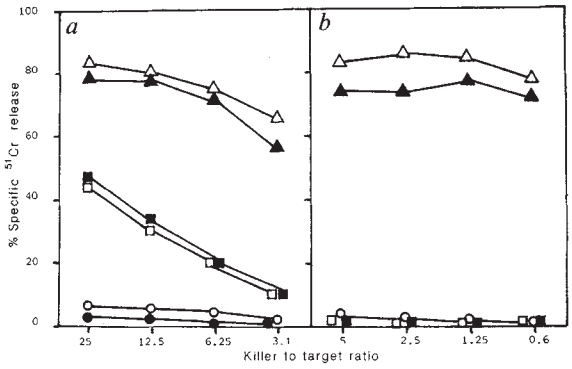


Fig. 4 CTL primed by full-length haemagglutinin recognize target cells expressing the signal-deleted molecule. *a*, Female CBA mice (3-6 months old) were primed intravenously with 100 haemagglutinin units of live A/PR/8/34 influenza virus. After 11 days spleen cells from primed animals were stimulated *in vitro* with A/PR/8/34 and collected for testing in the ^{51}Cr release assay as described in Fig. 3. Target cells used in this assay were L cells transfected with the Class I gene D^b (ref. 2). These cells were chosen so that the D^b-restricted nucleoprotein-specific CTL clone 11 could be used as an antigen specificity control. *b*, CTL clone 11 was isolated from a C57BL/10 mouse primed with influenza E61-13-H17 and restimulated with influenza X31 as described[20]. Clone 11 recognizes a conserved epitope on the nucleoprotein from human type A influenza viruses isolated since 1934 (J. Bastin, unpublished). Target cells were labelled with ^{51}Cr and infected with either recombinant vaccinia (5 PFU per cell) or influenza (200 μl of infectious allantoic fluid per 10^7 cells) for 1.5 h, washed twice, resuspended at 10^6 cells per ml in RPMI/10 for 3 h at 37 °C, washed twice and then used in the ^{51}Cr release assay as described[2]. Target cells were infected with; ■, H1-VAC; □, L⁻H1-VAC; ●, H3-VAC; ▲, A/PR/8/34 influenza (H1 haemagglutinin); △, X31 influenza (H3 haemagglutinin); ○, uninfected.

the appropriate antibodies could not detect HA on the surface of L⁻H1-VAC infected cells. Target cells infected with L⁻H1-VAC were not killed by a nucleoprotein specific T-cell clone, showing that recognition of the mutant HA was antigen specific. These experiments detect only CTL restricted through class I molecules of the major histocompatibility complex as the L cells used as targets do not express class II molecules.

These results are related to previous findings with nucleoprotein[5,6]. They show that CTL recognition of the haemagglutinin molecule (a typical glycosylated integral membrane protein[14]) does not depend on the presence of its signal sequence. Our findings suggest that an alternative route exists for the expression of HA at the surface of the cell in a form that is recognized by CTL. The lack of antibody binding to the L⁻H1-VAC-infected cell surface implies that the molecules recognized by CTL have disrupted tertiary structure. The pulse-chase experiment (Fig. 1) is consistent with this resulting from degradation of the mutant HA within four hours of its synthesis.

This conclusion is consistent with other recent published work. Wabuke-Bunoti *et al.* have shown that some HA-specific CTL can be induced *in vitro* with short synthetic peptides corresponding to regions of the H2 subtype haemagglutinin sequence[15]. Similarly, Yamada *et al.*[16] have described a recombinant protein produced in *Escherichia coli* consisting of the N-terminal 81 amino acids of the NS-1 protein fused to the C-terminal 210 amino acids of H1 haemagglutinin that was able to sensitize target cells for lysis by HA-specific CTL. We suggest, by analogy with the effect of denaturation on the recognition of proteins by T helper cells[17,18], that the tertiary structure of the HA fusion protein was disrupted enough to allow CTL recognition to occur.

Finally, Morrison *et al.*[19] have described Class I-restricted CTL that selectively recognized target cells in which new HA protein synthesis had occurred. In contrast their Class II-restricted CTL only recognized target cells that had been exposed to soluble HA in the surrounding medium. Our results argue that these findings can not be explained simply by the requirement for Class I-restricted CTL to recognize native HA inserted in the plasma membrane. We suggest that two pathways for presenting antigens to T cells exist, one dealing with proteins recently synthesized in the cytoplasm of an infected cell, the other concerned with endocytosed soluble proteins.

We thank G. Smith and B. Moss for the gifts of plasmid pGS62, recombinant vaccinia viruses H1-VAC and NP-VAC, and for initial help with vaccinia virus methodology. We are grateful to W. Gerhard, R. Webster and G. Schild for providing antibodies H36-18-2, 469/6 and R401 respectively; to N. Carter for analysing samples on the Ortho Cytofluorograf; to Claire Edmans for secretarial help and to the MRC for support.

Received 11 August; accepted 9 October 1986.

1. Zweerink, H. J., Askonas, B. A., Millican, D., Courtneidge, S. A. & Skehel, J. J. *Eur. J. Immun.* **7**, 630-635 (1977).
2. Townsend, A. R. M., McMichael, A. J., Carter, N. P., Huddleston, J. A. & Brownlee, G. G. *Cell* **39**, 13-25 (1984).
3. Yewdell, J. W., Bennink, J. R., Smith, G. L. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1785-1789 (1985).
4. Braciale, T. J., Braciale, V. L., Henkel, T. J., Sambrook, J. & Gething, M. J. *J. exp. Med.* **159**, 341-354 (1984).
5. Townsend, A. R. M., Gotch, F. M. & Davey, J. *Cell* **42**, 457-467 (1985).
6. Townsend, A. R. M. *et al. Cell* **44**, 959-968 (1986).
7. Gething, M. & Sambrook, J. *Nature* **300**, 598-603 (1982).
8. Gething, M. & Sambrook, J. in *Genetics of Influenza Viruses* 169-191 (Springer, New York, 1983).
9. Zoller, M. J. & Smith, M. *Meth. Enzym.* **100**, 468-500 (1983).
10. Smith, G. L., Murphy, B. R. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7155-7159 (1983).
11. Bennink, J. R., Yewdell, J. W., Smith, G. L., Moller, C. & Moss, B. *Nature* **311**, 578-579 (1984).
12. Bennink, J. R., Yewdell, J. W., Smith, G. L. & Moss, B. *J. Virol.* **57**, 786-791 (1986).
13. Winter, G., Fields, S. & Brownlee, G. G. *Nature* **292**, 72-75 (1981).
14. Sabatini, D., Kreibich, G., Morimoto, T. & Adesnik, M. *J. Cell Biol.* **92**, 1-22 (1982).
15. Wabuke-Bunoti, M. A. N., Taku, A., Fan, D., Kent, S. & Webster, R. G. *J. Immun.* **133**, 2194-2201 (1984).
16. Yamada, A., Young, J. F. & Ennis, F. A. *J. exp. Med.* **162**, 1720-1725 (1985).
17. Streicher, H. Z., Berkower, I. J., Busch, M., Gurd, F. R. N. & Berzofski, J. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6831-6835 (1984).
18. Allen, P. M. & Unanue, E. R. *J. Immun.* **132**, 1077-1079 (1984).
19. Morrison, L. A., Lukacher, A. E., Braciale, V. L., Fan, D. & Braciale, T. J. *J. exp. Med.* **163**, 903-921 (1986).
20. Townsend, A. R. M. & Skehel, J. J. *J. exp. Med.* **160**, 552-563 (1984).

H–2-restricted cytolytic T cells specific for HLA can recognize a synthetic HLA peptide

Janet L. Maryanski*, Pietro Pala*, Giampietro Corradin†, Bertrand R. Jordan‡ & Jean-Charles Cerottini*

* Ludwig Institute for Cancer Research, Lausanne Branch, 1066 Epalinges, Switzerland
† Institute of Biochemistry, University of Lausanne, 1066 Epalinges, Switzerland
‡ Centre d'Immunologie INSERM-CNRS de Marseille-Luminy, Case 906, Marseille Cedex 9, France

It is generally accepted that T lymphocytes recognize antigens in the context of molecules encoded by genes in the major histocompatibility complex (MHC). MHC class II-restricted T cells usually recognize degraded or denatured rather than native forms of antigen on the surface of class II-bearing antigen presenting cells (reviewed in refs 1, 2). It has recently been shown[3] that short synthetic peptides corresponding to mapped antigenic sites of the influenza nucleoprotein (NP) can render uninfected target cells susceptible to lysis by NP-specific class I-restricted cytolytic T cells (CTL). These and earlier experiments[4] that showed specific recognition of NP deletion mutant transfectants suggest that class I-restricted recognition might also involve processed antigenic fragments. One important issue arising from these studies is whether the model applies not only to viral proteins that are expressed internally (such as NP) but also to antigens normally expressed as integral membrane proteins at the cell surface. We have recently isolated class I-restricted mouse CTL clones that recognize class I gene products of the human MHC (HLA) as antigens in mouse cell HLA-transfectants[5]. Here we show that these anti-HLA CTL can lyse HLA-negative syngeneic mouse cells in the presence of a synthetic HLA peptide. These results suggest that the model applies generally.

CTL clones were isolated as described elsewhere[5] from DBA/2 (H-2^d) mice immunized with syngeneic cells transfected with cloned HLA class I genes. A high efficiency transfection recipient[6] selected from mouse cell line P815 was used for DNA transfection. Anti-CW3 CTL clone 10.1 lysed P815 cells transfected with HLA-CW3, but not HLA-A3, -A24 or -B7 (ref. 5 and J.L.M., unpublished results). Recognition by this and other DBA/2 CTL clones directed against P815-HLA transfectants was shown to be restricted to H-$2K^d$ (ref. 5). The antigenic determinant recognized by this CTL clone was mapped[7] to the second external (α_2) domain of HLA-CW3. Based on the amino-acid sequence of the α_2 domain of HLA-CW3 compared to the corresponding domain of HLA-A3, -A24 and -B7 and that of endogenous H-2^d class I molecules, we prepared a series of four synthetic peptides corresponding to 16-amino-acid segments of HLA-CW3 (Fig. 1).

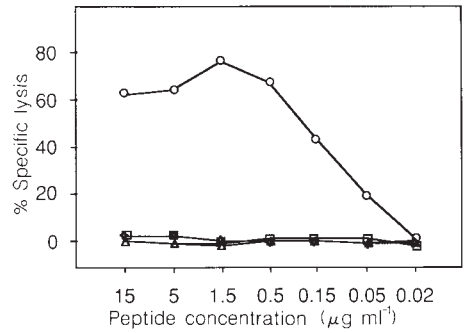


Fig. 2 Lysis of P815 target cells in the presence of peptide CW3(171–186) by anti-CW3 CTL clone 10.1. Synthetic peptides corresponding to HLA-CW3 amino acids 87–102 (△), 106–121 (◇), 151–166 (□) and 171–186 (○) were made according to the Merrifield method[19] as modified in ref. 20 using base-labile α-*N*-fluorenylmethoxycarbonyl protection. The peptides were cleaved from the resin with 90% aqueous trifluoroacetic acid, precipitated by addition of ether, redissolved in 10–20% acetic acid, chromatographed on a Sephadex G-25 column and eluted with 10–20% acetic acid. Peptides 87–102, 106–121 and 151–166 were then purified by reverse phase HPLC using a Permacoat C-6 column and elution with 0.1 M $NaClO_4$ in 0.1% H_3PO_4 buffer of *p*H 2 and a linear gradient of acetonitrile. Peptide 171–186 was purified by cation exchange HPLC using a Beckman Spherogel-TSK IEX 5.5 cm column and a linear gradient of NaCl in 0.01 M sodium phosphate buffer of *p*H 7.0. The peptides were lyophilized and dissolved in 0.7% sodium bicarbonate buffer. The HPLC-purified peptides gave the correct theoretical values of amino-acid composition and were >90% pure as judged by analytical HPLC. The purified peptides were diluted in Dulbecco's modified Eagle's medium (DMEM) containing 5% fetal bovine serum (FBS) and 100-μl volumes were added to wells of V-bottom microtitre plates. Cells (10^6) from P815-tk^+ clone 444/A1.1 (ref. 5) were labelled with 150 μCi sodium [^{51}Cr]chromate for 1 h at 37 °C and washed 3 times. Labelled P815 target cells (2×10^3 in 50-μl volumes) were added to wells containing peptides. 6×10^3 cells from anti-CW3 CTL clone 10.1 (ref. 5) in 50-μl volumes were then added to give an effector to target ratio of 3:1. After 4 h incubation at 37 °C the plates were centrifuged and 100 μl of supernatant was removed for counting. The % specific lysis was calculated as: (experimental − spontaneous release)/(total − spontaneous release) × 100. Spontaneous release in the absence of CTL or peptides was 14% of total counts. There was no effect on spontaneous release by any of the peptides. The specific lysis in the absence of peptides was 1%.

The CW3-specific CTL clone 10.1 lysed P815 cells incubated with peptide CW3 (171–186) very efficiently (Fig. 2). The minimum peptide concentration required for maximum lysis in this assay was 1.5 μg ml^{-1} (0.8 μM). The other three CW3 peptides tested were negative up to a concentration of 15 μg ml^{-1}. Another independently-derived H-$2K^d$-restricted DBA/2 anti-CW3 CTL clone (P1.1) apparently recognized the same HLA-

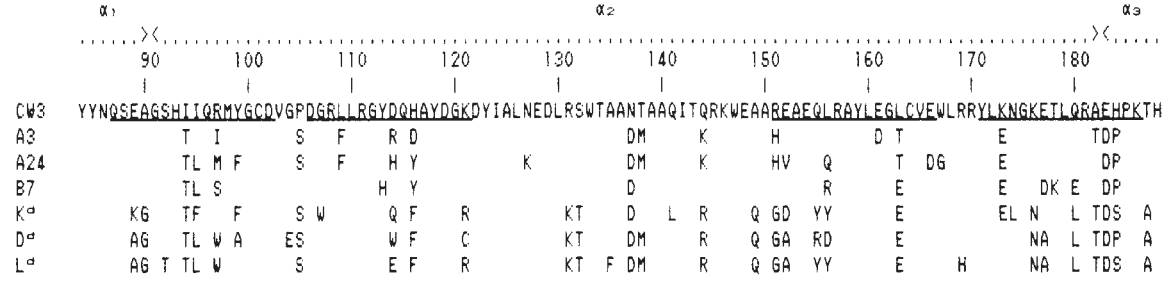


Fig. 1 Comparison of the protein sequence of HLA-CW3[10], -A3[11], -A24[12], and -B7[13], and H-$2K^d$ (ref. 14), -D^d (ref. 15), and L^d (ref. 16) from amino acid 83 to 188. Amino acid 181 of the HLA-CW3 sequence, erroneously given as G in ref. 10, has been corrected to R[17]. For H-$2L^d$, amino acids 121, 139 and 140 are given as C, S, and S, respectively in ref. 18. The four peptide sequences prepared synthetically are underlined.

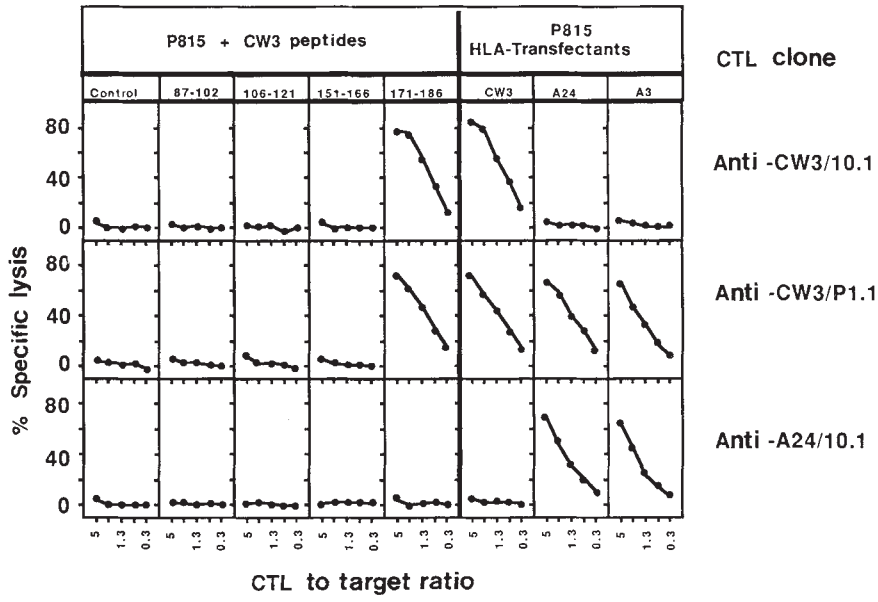


Fig. 3 Specificity of HLA-CW3 peptide recognition. Peptides CW3(171-186) and CW3(87-102) (100 μl at 10 μg ml^{-1} in medium containing 5% FBS) were added to conical microtitre wells. Cells (10^6) from P815 clones[5] transfected with the thymidine kinase (*tk*) gene only (clone 444/A1.1), with *tk* and HLA-CW3 genes (444/C9.3), with *tk* and HLA-A24 genes (452/D1), or with *tk* and HLA-A3 genes (504/2.1.4) were labelled with 150 μCi sodium [^{51}Cr]chromate for 1 h at 37 °C and washed 3 times. Labelled target cells (2×10^3 in 50-μl volumes) were added to wells that contained medium or peptides. Anti-CW3 CTL clones 10.1 and P1.1 were derived in independent experiments from DBA/2 mice immunized with P815 HLA-CW3 transfectant clone 444/C9.3, and anti-A24 CTL clone 10.1 was derived from DBA/2 mice immunized with P815 HLA-A24 transfectant cells[5]. Serial twofold dilutions of CTL in 50-4μl volumes were added to give the indicated CTL to target ratios and a final peptide concentration of 5 μg ml^{-1}. After 4 h at 37 °C the plates were centrifuged and 100 μl supernatant was removed for counting. Specific lysis was calculated as in Fig. 2.

CW3 synthetic peptide as clone 10.1 (Fig. 3). The fine specificity of the two CTL clones was not exactly identical, however. Clone P1.1 lysed P815-A24 and -A3 targets in addition to CW3 targets, unlike clone 10.1 that lysed only CW3 targets (Fig. 3 and ref. 5). Neither lysed HLA-negative P815-*tk*$^+$ cells. A third H-2K^d-restricted DBA/2 anti-HLA CTL clone that lysed HLA-A24 and -A3, but not -CW3 transfectants failed to react with any of the HLA-CW3 peptides, at concentrations of up to 30 μg ml^{-1} (Fig. 3 and J.L.M., unpublished results). Thus peptide recognition by these CTL clones (as measured by cytolytic activity) is specific, and antigenic sites for two different anti-CW3 CTL clones may be located within the same 16-amino-acid sequence.

These results clearly agree with those in ref. 3 showing specific cytolytic activity of anti-NP CTL in the presence of synthetic NP peptides. The reconstitution of specific target cell recognition with synthetic peptides strongly supports the hypothesis that class I-restricted T cells, like those restricted to class II antigens, recognize degraded or denatured antigenic fragments rather than intact antigen. It is noteworthy that HLA antigens, unlike the influenza NP, are integral cell membrane glycoproteins that can be easily detected on the cell surface with specific antibody[5,8]. Our results suggest, therefore, that recognition of antigenic fragments by class I-restricted T cells may be a general feature not limited to antigens that are virus-encoded and/or expressed primarily intracellularly.

Previous studies have shown that HLA class I gene products expressed in P815 mouse cells are involved in at least three categories of T-cell recognition. These include alloantigens recognized by alloreactive human CTL[8], HLA class I restriction elements recognized by human anti-viral CTL[9], and nominal antigens recognized by class I-restricted mouse CTL[5]. We have shown in this paper that the nominal HLA-CW3 antigen for two different class I-restricted CTL clones can be defined by a synthetic 16-amino-acid peptide. We have previously observed that lysis of P815-HLA transfectants by alloreactive human CTL but not by class I-restricted mouse anti-HLA CTL can be blocked by anti-HLA antibodies[5,8]. Together the results suggest that these human and mouse CTL recognize different forms of HLA that are presumably encoded by the same gene. It is possible that a membrane bound form of HLA in P815 HLA transfectants is recognized by alloreactive and by HLA-restricted antiviral human CTL, whereas a processed fragment of HLA is recognized as a nominal antigen by class I-restricted mouse CTL. The native and degraded forms of antigen may coexist on the cell surface. We believe P815-HLA transfectants will constitute an ideal system to investigate the nature of antigen complexes recognized by T cells.

We thank K. Mühlethaler and C. Vonnez for expert technical assistance and Dr H. R. MacDonald for helpful comments.

Received 15 September; accepted 10 October 1986.

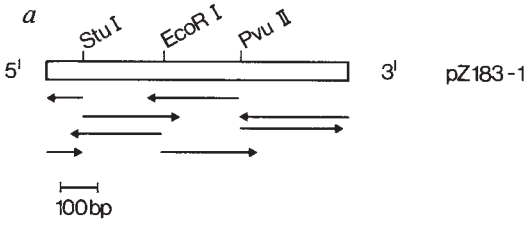
1. Unanue, E. R. *A. Rev. Immun.* **2**, 395-428 (1984).
2. Grey, H. M. & Chestnut, R. *Immun. Today* **6**, 101-106 (1985).
3. Townsend, A. R. M. *et al. Cell* **44**, 959-968 (1986).
4. Townsend, A. R. M., Gotch, F. M. & Davey, J. *Cell* **42**, 457-467 (1985).
5. Maryanski, J. L., Accolla, R. S. & Jordan, B. *J. Immun.* **136**, 4340-4347 (1986).
6. Van Pel, A., De Plaen, E. & Boon, T. *Somatic Cell molec. Genet.* **11**, 467-475 (1985).
7. Pala, P., *et al.* in preparation.
8. Maryanski, J. L. *et al. Eur. J. Immun.* **15**, 1111-1117 (1985).
9. Gomard, E. *et al. Nature* **319**, 153-154 (1986).
10. Sodoyer, R. *et al. EMBO J.* **3**, 879-885 (1984).
11. Strachan, T., Sodoyer, R., Damotte, M. & Jordan, B. R. *EMBO J.* **3**, 885-894 (1984).
12. N'Guyen, C., Sodoyer, R., Trucy, J., Strachan, T. & Jordan, B. R. *Immunogenetics* **21**, 479-489 (1985).
13. Orr, H. T., Lopez de Castro, J. A., Parham, P., Ploegh, H. L. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4395-4399 (1979).
14. Kvist, S., Roberts, L. & Dobberstein, B. *EMBO J.* **2**, 245-254 (1983).
15. Sher, B. T., Nairn, R., Coligan, J. E. & Hood, L. E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1175-1179 (1985).
16. Evans, G. A., Margulies, D. H., Camerini-Otero, R. D., Ozato, K. & Seidman, J. F. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1994-1998 (1982).
17. Jordan, B. R., Lemmonier, F. A., Caillol, D. H. & Trucy, J. *Immunogenetics* **18**, 165-171 (1983).
18. Moore, K. W., Sher, B. T., Sun, Y. H., Eakle, K. A. & Hood, L. *Science* **215**, 679-682 (1982).
19. Marglin, A. & Merrifield, R. B. *A. Rev. Biochem.* **39**, 841-866 (1970).
20. Atherton, E., Logan, C. J. & Sheppard, R. C. *J. chem. Soc. Perkin* **I**, 538-546 (1981).

λ5, a new light-chain-related locus selectively expressed in pre-B lymphocytes

Nobuo Sakaguchi & Fritz Melchers

Basel Institute for Immunology, Grenzacherstrasse 487, Postfach, CH-4005 Basel, Switzerland

The development from stem cells to pre-B cells, B lymphocytes and, finally, plasma cells and memory cells proceeds through various stages which have been defined by the genomic context in which immunoglobulin (Ig) heavy (H) and light (L) chain gene segments are found, as well as by their state of expression[1-7]. They have also been identified by surface marker analysis[5,8,9] and susceptibility to various stimuli regulating growth and differentiation[10]. We have searched for genes that are expressed at given stages in the B-lymphocyte development pathway and which might function to control this development at various stages. A com-



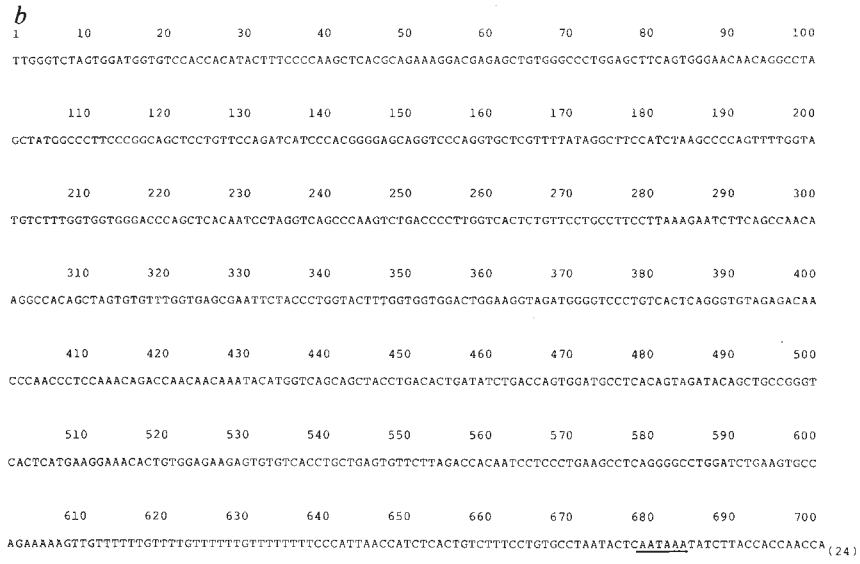


Fig. 1 Strategy for the sequencing and nucleotide sequence of the cDNA clone pZ183-1. The pZ183 cDNA clone was obtained as described elsewhere[11]. ^{32}P-labelled insert DNA was used to screen a pUC-13 cDNA library constructed from total poly(A)$^+$ RNA of 70Z/3 cells by the tailing method[25]. The restriction enzyme sites and the sequences are shown in *a*. The arrows indicate the direction and length of sequence determined by subcloning fragments into M13 phage vector and sequencing by the dideoxy-chain termination method[26]. The nucleotide sequence of the cDNA clone pZ183-1 is shown in *b*. The 5′ to 3′ orientation was deduced from the location of the poly(A) tail and the poly(A) attachment site (underlined). Nucleotide positions are numbered from the most 5′ position of the insert cDNA fragment.

plementary DNA sequence called pZ183 was found in a library constructed from messenger RNA of the murine pre-B lymphoma cell line 70Z/3 which is selectively expressed in pre-B cells[11]. Here we report the nucleotide sequence of a cDNA clone (pZ183-1) containing 0.7 kilobases (kb) of the pZ183 gene. Part of this sequence shows strong homology to constant (*C*) and joining (*J*) region sequences of λ_1 L chains. Our findings define a new immunoglobulin L-chain-related locus, which we call λ5, that is selectively transcribed in pre-B lymphocytes.

During the somatic development of lymphocytes from precursor cells, gene segments in immunoglobulin gene loci are rearranged to form a complete variable (*V*) region exon which results in the translation into full immunoglobulin chains containing V- and C-region domains[12,13]. In the B-cell lineage rearrangement of heavy-chain diversity (D_H) segments to J_H segments precedes that of V_H to D_HJ_H segments[3,14,15]. V_κ is then rearranged to J_κ (refs 2, 16, 17) and, finally, V_λ to J_λ (ref. 18). The different steps of H, κ *L* and λ L rearrangments define different cellular stages of pre-B-cell development.

In our search for genes that are expressed at a given stage of pre-B-cell development, we had constructed a cDNA library from mRNA of (uninduced) 70Z/3 which was subtracted by a pool of mRNA molecules from a T-cell hybridoma[11]. We cloned the pZ183 cDNA which we found expressed selectively in pre-B cells at low to medium abundance[11]. It was present in 1 of 5,000 in the unsubtracted library and in 1 of 350 cDNA clones in a subtracted library from the same source of poly(A)$^+$ RNA of 70Z/3. Mature B cells, plasma cells, T cells, macrophages and a variety of cells from other lineages and tissues did not express detectable levels of the corresponding gene.

The cDNA clone with the longest (0.7 kb) insert, called pZ183-1, was obtained from a cDNA library constructed in the pUC-13 vector from a poly(A)$^+$ mRNA of uninduced 70Z/3 and sequenced as outlined in Fig. 1*a*. The nucleotide sequence is shown in Fig. 1*b*. Next to the poly(A) tail lies a poly(A) attachment site which orients the pZ183-1 sequence from 5′ to 3′. The three longest open reading frames deduced from the cDNA sequence were considered as possible candidates for the polypeptide sequence: the longest frame between bases 1 and 553 (184 residues), the second between bases 9 and 281 (93 residues) and the third between bases 183 and 462 (93 residues).

We then searched a DNA sequence data bank[19] for possible evolutionary relationships of the pZ183-1 sequence to known DNA sequences. Strong homology was detected for positions

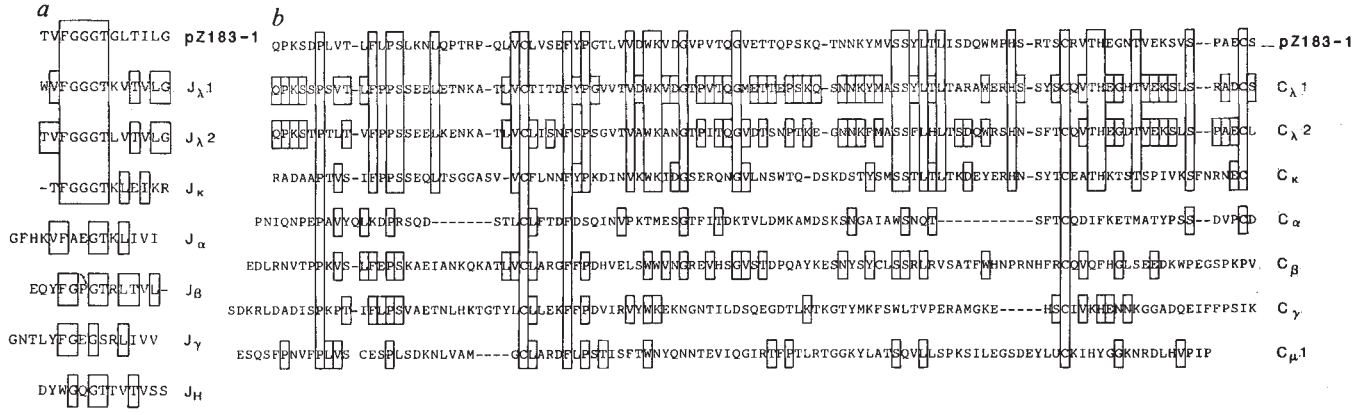


Fig. 2 Possible amino acid sequence of the *J*-like (*a*) and *C*-like (*b*) portions of pZ183-1 deduced from the nucleotide sequence in Fig. 1*b* in frame 1 and alignment with the corresponding sequences of λ1 L (refs 21, 27), λ2 L (refs 21, 27), κ L (ref. 21), J_H and $C_\mu1$ (ref. 21) and J_α $C_\alpha1$, J_β C_β and J_γ C_γ of the α, β and γ chains of the T-cell receptor[22,24,28]. See Table 1 for nucleotide positions corresponding to first and last deduced amino acid of the *J*-like and *C*-like sequences of pZ183-1. Those residues common between pZ183-1 and any other sequence are boxed. Residues conserved in all sequences or conserved in light chains are boxed together.

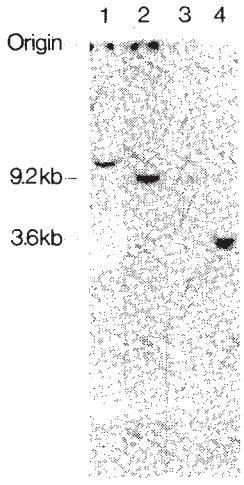


Fig. 3 Southern blot analyses of DNA from 70Z/3 cells. DNA was digested with the restriction enzyme *Hin*dIII, electrophoresed through a 1.0% agarose gel, blotted onto nitrocellulose filters and hybridized with the ^{32}P-labelled insert of the clone pZ183-1 (lane 1). Hybridization was carried out at 42 °C in 50% formamide and 5×SSC. The filters were washed at 65 °C in 0.2×SSC. After exposure to X-ray film, the same filter was washed in H_2O 0.1% SDS at 95 °C for 10 min and used repeatedly to probe successively with $C_\lambda 1$ [*Xho*I–*Pst*I fragment of pAB1-1 (ref. 27) (lane 2), both pZ183-1 and $C_\lambda 1$ (lane 3)] and $C_\lambda 2$ (450-bp *Sac* I fragment[23] (lane 4).

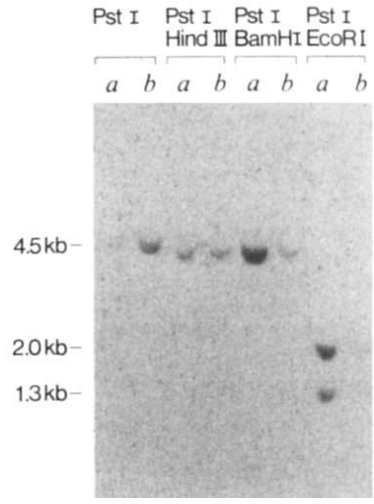


Fig. 4 Southern blot analyses of DNA from liver of (C57BL/6J × DBA/2)F_1 mice (*a*) and from 70Z/3 pre-B cells (*b*). DNA was digested with the indicated enzymes. The filter was hybridized with ^{32}P-labelled insert DNA of cDNA clone pZ183-1 under conditions described in Fig. 3 legend.

240–554 to the *C*-region nucleotide sequence of λ1 L chains. Equally strong homology was found for the adjacent 5′ positions 201–239 to the *J* segment sequence of λ1 (Table 1). The reading frame of the λ1 L-chain *J*+*C* nucleotide sequence was then imposed on the pZ183-1 sequence as a reading frame. This was found to be reading frame 1 (starting with nucleotide 1) of the sequence shown in Fig. 1*b*. Comparison with the λ1 nucleotide sequence shows that the first stop codon in frame 1 of the pZ183-1 sequence occurs at the same position as the first (functional) stop codon of λ1 L chains. Reading frame 1 is therefore the most likely candidate for a translatable sequence.

The amino acid sequence encoded by the *J*- and *C*-like regions of the pZ183-1 gene deduced from reading frame 1 is shown in Fig. 2 and compared with corresponding sequences of immunoglobulin and T-cell receptor proteins. Again, the strongest homology is detected with the *J* segment and *C* domain of λ1 L chains (Table 1).

Table 1 Homologies of pZ183-1 sequences with immunoglobulin light-chain genes and proteins

Sequence position of pZ183-1	L chain	Ref.	Homology with Nucleotide sequence (%)*	Homology with deduced amino acid sequence (%)†
	$C_\lambda 1$	21, 22	77.9	66.7
	$C_\lambda 2$	21, 22	70.2	61.1
240–554	$C_\lambda 3$	22, 23	70.9	61.1
	$C_\lambda 4$	23	73.5	—
	C_κ	22, 24	22.5	30.8
	$J_\lambda 1$	21, 22	67.5	66.0
	$J_\lambda 2$	21, 22	70.2	75.0
201–239	$J_\lambda 3$	22, 23	67.6	50.0
	$J_\lambda 4$	23	70.2	—
	$J_1\kappa$	22,24	59.4	58.0

* Nucleotide sequence comparisons were done by a computer program and are shown as per cent matches per length.

† Alignment of the deduced amino acid sequence of pZ183-1 with mouse λ1, λ2, λ3 and λ4 L chains, as well as with κ L chains was done to maximize identities and structural homologies with pZ183-1, paying special attention to the alignment of certain key conserved residues.

Furthermore, pZ183-1 shows cysteine residues at the consensus positions characteristic of all immunoglobulin domains, and has the consensus sequence WVFGGGTKVTVLG characteristically encoded by *J* segments of the immunoglobulin genes. These results suggest that the 3′ portion of the pZ183-1 cDNA clone belongs to a λ L-chain-related gene locus which we call λ5 ($J_\lambda 5$, $C_\lambda 5$).

In contrast, the 5′ sequence adjacent to the $J_\lambda 1$-like sequence of pZ183-1 shows no strong homology to $V_\lambda 1$, any other known *V* segment or any other gene known to us. It does not contain a cysteine at the position where a consensus cysteine residue would be expected in V-region domains of immunoglobulins. The 5′ sequence adjacent to the *J*-like sequence of pZ183-1 also reveals no homologies to intron nucleotide sequences either 5′ of $J_\lambda 1$ or 3′ of $V_\lambda 1$. The segment cDNA clone is apparently too short to contain an initiation codon. We therefore cannot rule out the possibility that the pZ183-1 cDNA is derived from a sterile transcript.

The gene is carried in the genome of 70Z/3 cells in a *Hin*dIII fragment of >9.2 kb (Fig. 3, lane 1). Southern blot analysis shows that it is different in context from $C_\lambda 1$ (Fig. 3, lane 2) and $C_\lambda 2$ (lane 4). The unknown 5′ sequences and the $J_\lambda 1$- and $C_\lambda 1$-like 3′ sequences of pZ183-1 appear to be on the 4.5-kb *Pst*I fragment. The gene is on one *Pst*I/*Bam*HI and one *Pst*I/*Hin*dIII genomic fragment, each 4.2 kb long. Double digestion with *Pst*I and *Eco*RI yields two fragments of 1.3 and 2.0 kb (Fig. 4). No differences in the lengths of restriction enzyme fragments of DNA from liver and from 70Z/3 cells have been detected (Fig. 4). This indicates that the 5′ and 3′ regions of the pZ183-1 gene are near each other, that is, on one 4.3-kb fragment in germline DNA. We have therefore not found any sign that they are rearranged during pre-B-cell development. This also makes it unlikely that a chromosomal translocation or viral insertion occurred in close proximity to the pZ183-1 sequences in either the chemically induced 70Z/3 pre-B cell or the Abelson virus-induced pre-B-cell lines in which the same gene is expressed with the same size transcript of 1.2 kb[11].

The size of pZ183-1 transcripts, 1.2 kb, predicts a protein of similar size to L chains. With its penultimate cysteine at the carboxy-terminal end it could form homodimers (for example,

L-chain dimers) or heterodimers. H chains could be partners in this heterodimer formation as soon as they are expressed in pre-B cells, and this may allow H chains to appear on the surface of pre-B cells, as previous experiments have indicated[9]. Such a surface deposition of the expressed H chains would significantly alter our present view of the development of the B-cell repertoire.

It had previously appeared that all the loci in mouse which encode antigen-binding immunoglobulin molecules of B lymphocytes had been identified, although considerable changes in the complexity of λ1 L-chain loci has been observed in wild mice[28]. However, our results suggest that immunoglobulin-related genes may exist which function only at an early stage in ontogeny and which so far have gone undetected.

We thank Ms Heidi Bächtold and Denise Richterich for technical assistance and Dr Klaus Karjalainen for critical reading of our manuscript. The Basel Institute for Immunology was founded and is supported by F. Hoffmann-La Roche & Co., Basel, Switzerland.

Received 10 July; accepted 11 September 1986.

1. Tonegawa, S. *Nature* **302**, 575-581 (1983).
2. Maki, R., Kearney, J., Paige, C. & Tonegawa, S. *Science* **209**, 1366-1369 (1982).
3. Alt, F. W., Rosenberg, N., Lewis, S., Thomas, E. & Baltimore, D. *Cell* **27**, 381-390 (1981).
4. Perry, R. P., Kelley, D. E., Coleclough, C. & Kearney, J. F. *Proc. natn. Acad. Sci. U.S.A.* **78**, 247-251 (1981).
5. Coffman, R. L. & Weissman, I. L. *Molec. cell. Immun.* **1**, 33-38 (1983).
6. Paige, C. J., Kincade, P. W. & Ralph, P. *J. Immun.* **121**, 641-647 (1978).
7. Reth, M. G., Ammirati, P., Jackson, S. & Alt, F. W. *Nature* **317**, 353-355 (1985).
8. McKearn, J. P., Baum, C. & Davie, J. M. *J. Immun.* **132**, 332-339 (1984).
9. Kincade, P. W., Lee, G., Watanabe, T., Sun, L. & Scheid, M. P. *J. Immun.* **126**, 2262-2266 (1981).
10. Melchers, F., Andersson, J. & Phillips, R. A. *Cold Spring Harb. Symp. quant. Biol.* **41**, 147-158 (1977).
11. Sakaguchi, N., Berger, C. N. & Melchers, F. *EMBO J.* **5**, 2139-2147 (1986).
12. Hozumi, N. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3628-3632 (1976).
13. Sakano, H., Maki, R., Kurosawa, Y., Roder, W. & Tonegawa, S. *Nature* **286**, 676-683 (1980).
14. Sugiyama, H. *et al. Nature* **303**, 812-815 (1983).
15. Alt, F. W. *et al. EMBO J.* **3**, 1209-1219 (1984).
16. Lewis, S., Rosenberg, N., Alt, F. & Baltimore, D. *Cell* **30**, 807-816 (1983).
17. Burrows, O. D., Beck-Engeser, G. B. & Wabl, M. R. *Nature* **306**, 243-246 (1983).
18. Hieter, P. A., Korsmeyer, S. J., Waldmann, T. A. & Leder, P. *Nature* **290**, 368-372 (1981).
19. *GenBank, Genetic Sequence Data Bank* Release 35.0 (1985).
20. Scott, C. L., Mushinski, J. F., Huppi, C., Weigert, M. & Potter, M. *Nature* **300**, 757-760 (1982).
21. Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. & Perry, H. in *Sequences of Proteins of Immunological Interest* (NIH, Bethesda, 1983).
22. Hedrock, S. M., Nielsen, E. A., Kavaler, J., Cohen, D. I. & Davis, M. M. *Nature* **308**, 253-258 (1984).
23. Weiss, S., Meyer, J. & Wabl, M. R. *Eur. J. Immun.* **15**, 765-768 (1985).
24. Saito, H. *et al. Nature* **309**, 757-762 (1984).
25. Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning. A Laboratory Manual*, 230-242 (Cold Spring Harbor Laboratory, New York, 1982).
26. Sanger, F., Nicklen, S. & Coulson, A. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
27. Bothwell, A. L. M. *et al. Nature* **298**, 380-383 (1982).
28. Saito, H. *et al. Nature* **312**, 36-40 (1984).

Long-range restriction map around the Duchenne muscular dystrophy gene

Margit Burmeister & Hans Lehrach

EMBL, Postfach 102209, Meyerhofstrasse, 6900 Heidelberg, FRG.

Duchenne muscular dystrophy (DMD, reviewed in ref. 1) is an X-linked recessive disease affecting about 1 in 4,000 newborn boys. As in many other inherited diseases, the biochemical basis of the condition is unkown, and as yet there is no effective treatment. Translocations, deletions and other mutations leading to the DMD phenotype are distributed over a chromosomal area of large, but unknown size[2,3]. Using pulsed-field gradient gel electrophoresis[4,5], we have now determined restriction maps of a major fraction of this area, covering two regions of three million basepairs in total, and used it to determine the position of several probes linked to DMD. The maps establish physical distances between structural changes associated with the DMD phenotype and provide evidence for a CpG-rich island proximal to the area containing translocations and deletions associated with the DMD phenotype.

Using either molecular or genetic mapping, translocations, deletions or other mutations causing the DMD phenotype have been found distributed over an unusually large area of the X chromosome[2,3]. Although more complex genetic arrangements are possible, the simplest explanation is a single, but very large gene responsible for DMD. Two probes closely linked to the DMD gene and probably contained within have been cloned using different approaches. Subclones of a 200-kilobase (kb) cluster cloned around probe pERT87(DX164)[6] are deleted in about 6.5% of all DMD patients[3], suggesting that parts of the DMD gene might be deleted. Probe pXJ1.1 (ref. 7) is also assumed to be at the DMD gene, since it is derived from the chromosomal breakpoint of an X-chromosomal translocation found in a female DMD patient. As in other female DMD patients[2], the disease is assumed to arise due to disruption of the DMD gene by the translocation. Although extensive chromosomal walking has been carried out around both probes (P. Ray and R. Worton, and A. Monaco and L. Kunkel, personal communications), the two cloned regions could not be joined.

In contrast to most other genes, the size of the region implicated in the DMD phenotype is therefore too large to be easily accessible to standard molecular techniques. Recently, however, a new method, pulsed-field gradient (PFG) gel electrophoresis[4,5], has been developed, allowing the analysis of DNA fragments of 50-1,500 kilobases (kb); this range is not covered either by molecular cloning or by classical genetic approaches. We therefore decided that the analysis of the DMD locus would be considerably facilitated by mapping the positions of cloned probes relative to each other. Since these probes are associated with structural changes leading to a DMD phenotype, the positions of mutations leading to DMD are determined and estimations of a minimal gene size are possible.

We have analysed five probes closely linked to and flanking the DMD gene using the PFG technique. In addition to probes pERT87 and pXJ1.1, which are very close to or at the DMD gene, the closest available polymorphic markers on the distal side, B24 (DXS67) and C7(DXS28)[8], about 5-15 cM from DMD[9], and L1-4 (DXS68)[10], a non-polymorphic marker localized between pERT87 and C7 (ref. 9), were included. Deletion analysis[9,11-13] indicates the probe order pXJ1.1/pERT87/L1-4/C7/B24, and maps the genes for glycerol kinase and adrenal hypoplasia between pERT87 and C7.

To establish a physical map in this region, DNA was cleaved with different restriction enzymes cutting rarely in the mammalian genome, separated on PFG gels, and transferred to nylon filters for analysis. Filters were hybridized successively with each of the five probes. Probes pERT87-30 (the most distal subclone of the pERT87-cluster) and pXJ1.1 detected common restriction fragments; B24 and C7 were similarly shown to be physically linked. To generate a restriction map, double digests with combinations of enzymes were performed and analysed (see Fig. 1). The results of this analysis are summarized in Table 1. In a manner analogous to conventional restriction mapping, the data were used to deduce restriction maps covering hundreds of kilobase pairs. The restriction maps constructed (see Fig. 2 and legend for details) show that probes pXJ1.1 and pERT87-30 are at least 100, but no more than 400 kb apart. This result also gives the distance between the X/21 DMD translocation[7] (of which pXJ1.1 is a junction fragment) and common DMD deletions (100 kb proximal of pERT87-30, around pERT87-8)[3] as maximally 300 kb (see Fig. 2). The size of the DMD gene is probably larger, however, since some mutations map distal to pERT87-8.

Overlapping clones which extend ~200 kb proximal to pERT87-30 have been isolated without joining the segment cloned around pXJ1.1 (ref. 3 and A. Monaco and L. Kunkel, personal communication); thus the maximum additional walking distance between the two cloned segments is now no more than 200 kb, and more likely to be much less.

B24 and C7 are the closest published markers distally flanking

the DMD gene[9]. They are also of particular interest, since deletions which produce a syndrome of adrenal hypoplasia, glycerol kinase deficiency and DMD, map proximal to these markers[11-13]. Our results show that these two probes lie within a distance of 250-700 kb from each other. We could not show physical linkage between these probes and the pERT87/pXJ1.1 cluster (Fig. 2, Table 1). In an attempt to detect physical linkage between the two clusters, probe L1-4, which had been localised between C7 and pERT87-30 by deletion analysis[9,11], was hybridized to the same filters. No common restriction fragments were

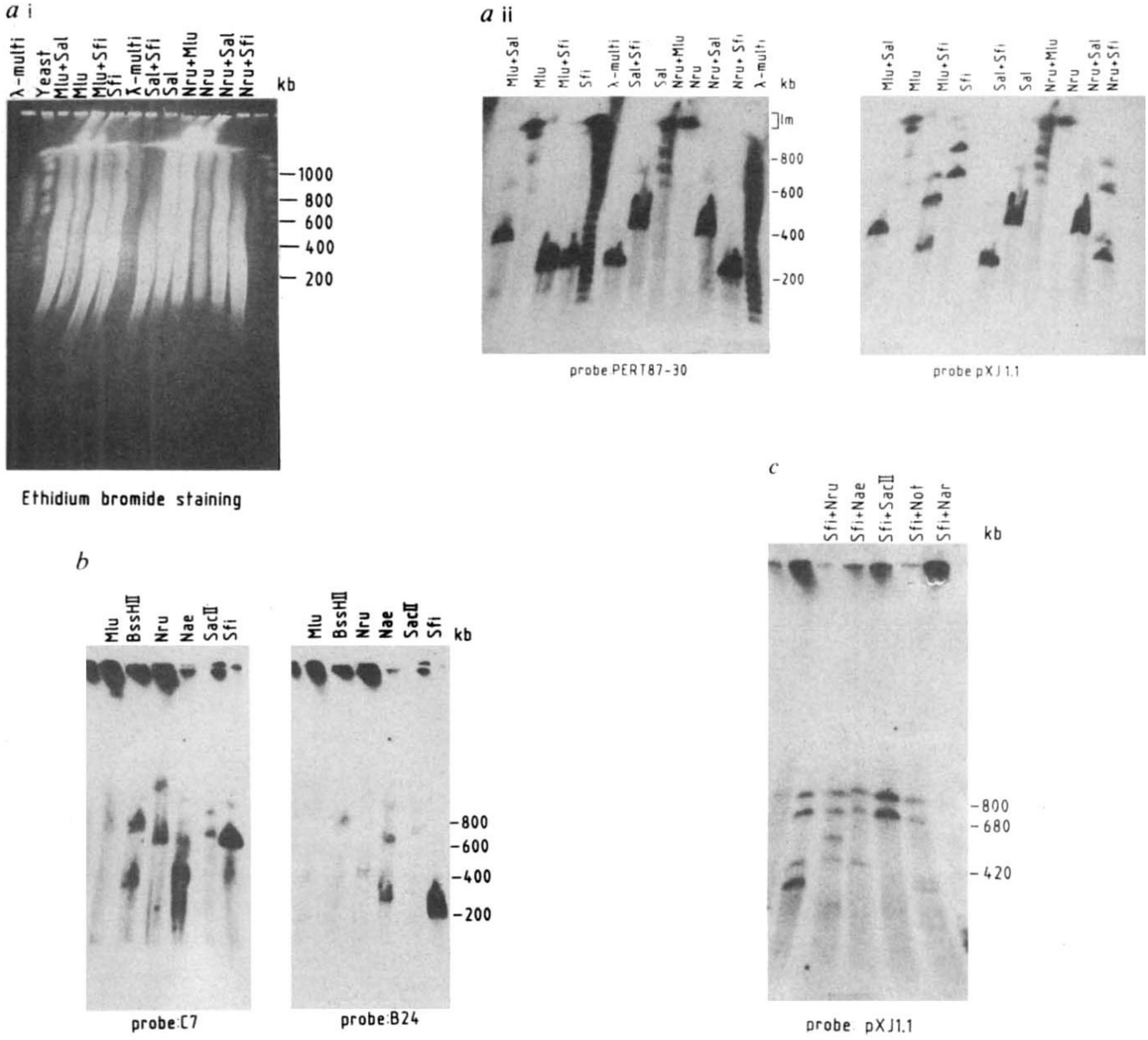


Fig. 1 Hybridizations to PFG gels. *a*, Linking probes pXJ1.1 and pERT87-30; *b*, common bands detected by probes B24 and C7; *c*, indications for an CpG-rich island near probe pXJ1.1. *a*, Ethidium bromide staining and autoradiograms of a gel run for 38 h at 15 °C with alternating switching of 48 to 90 seconds; These conditions give good resolution in the 200-600 kb range, but no resolution above about 900 kb. Size markers were both yeast chromosomes[4] and λ EMBL3 (ref. 18) polymers (λ-multi), which were generated by annealing the *cos* ends in agarose blocks. The unit size of EMBL3 is 43.2 kb. The autoradiograph at left shows the hybridization of λ DNA in addition to the probe used to visualize the size markers. The region of limiting mobility, lm, is the position of DNA fragments bigger than about 900 kb. *b*, This gel was run for 39 h at 9 °C with alternating switching times of 30 to 130 s, resolving also DNA fragments above 1,100 kb. Enzyme digests and probes are as indicated in the figure. Yeast chromosomes run on the same gel were used as size markers. c, Double digests of DNA with *Sfi*I and other enzymes as indicated were run on a PFG gel for 40 h at about 20 °C with alternating pulses of 30 and 130 s. *Sfi*I + *Not*I is equivalent to *Sfi*I alone. A common 390-kb fragment is detected with four enzymes in double digests with *Sfi*I. (*Sfi*I/*Bss*HII digests result in a similar fragment not shown in this figure). A 390-kb fragment is also observed when DNA is digested with *Nae*I alone. Therefore, the 390-kb fragment seen in the *Sfi*I + *Nae*I lane does not argue for the localization of an NaeI site close to the *Bss*HII/*Sac*II/*Nar*I/*Nru*I cluster.

Methods. Preparation of DNA for PFG was essentially as described[19]. Exact conditions for DNA preparation, running conditions and blotting will be published elsewhere[20]. A detailed protocol is available on request. The cell line LCL127 (48, XXXX, a gift from U. Francke, Yale) was used as a source for DNA in all cases. The source of the DNA is important because of observed methylation differences. Many sites seen in this cell line were not detectably cut in male blood cell DNA. Restriction enzymes were from New England Biolabs and were used as recommended by the manufacturer. Pulsed field gel electrophoresis[4] was carried out in a slightly modified box. The field was inhomogenous in both directions, the voltage gradient 10 V cm^{-1}. Pulse times are as indicated above. After staining and photography, gels were alkaline-blotted onto Gene Screen nylon membranes. Inserts were isolated from low melting point agarose gels labelled by random oligo-priming[21] to a specific activity of around $1\text{-}4\times10^9$ d.p.m. μg^{-1} DNA. Hybridisations were carried out as described[22] for at least 24 h. Exposure was for 1 to 7 days at −70 °C with Kodac-X-Omat films preflashed once. Before reprobing, filters were stripped with 2 mM Tris/HCl *p*H 8, 0.1% sodium dodecyl sulfate at 80 °C. When the strongest band from the previous hybridization was not visible after reprobing, stripping was considered complete.

Table 1 Summary of hybridization results

Enzymes	B24	C7	Probes L1-4	pERT87-30	pXJ1.1
*Sfi*I	200	650 (1000)	220 350	230	680 800
*Mlu*I	—	—	—	630 850 >950	630 850 >950
*Bss*HII	(350) 750	400 750	(850)	—	—
*Sac*II	(700) (900)	(700) (900)	(800)	(710)	(710)
*Sal*I	ND	ND	300	400 (650)	400 (650)
*Nru*I	450	700 1100	—	—	—
*Sfi*I + *Mlu*I	200	(500) 650	220 350	230 (150)	300 (500) 540 680
*Sfi*I + *Bss*HII	200	400 520 580 650	220 350	230	390 680 800
*Sfi*I + *Sac*II	200	500 650	220 350	230	390 680 800
SfiI + *Sal*I	ND	ND	220	220 (450)	200 (450)
*Sfi*I + *Nru*I	200	420 650	220 350	230	300 390 680 800
*Sac*II + *Bss*HII	ND	400 700 900	ND	ND	(710)
*Sac*II + *Nru*I	ND	250 450	ND	ND	ND
*Sac*II + *Mlu*I	ND	(300) 700	ND	ND	ND
*Sal*I + *Nru*I	ND	ND	ND	400 (650)	400 (650)
*Sal*I + *Mlu*I	ND	ND	300	(120) 400 630	(320) 400 (480) 630
*Mlu*I + *Nru*I	ND	ND	ND	630 700 850 900 >950	(360) 730 700 850 900 >950
*Nru*I + *Bss*HII	ND	350 400 750	ND	ND	ND

Sizes of bands are in kilobase pairs. A dash indicates that no band was observed; ND, not determined. To confirm identity of bands recognized by two probes, several gels were run using different pulse times giving optimal resolution for different size ranges, minimizing the accidental comigration of non identical bands. For the construction of the map, we also considered the intensity of bands relative to the strong *Sfi*I-bands. Strong bands (>30% of the intensity of *Sfi*I, considered cut to completion) are underlined and very weak bands (<5% the intensity of *Sfi*I) are in brackets. Note that weak bands can arise from a site which is cut to completion, when the second site is incompletely cut. The partial bands produced by *Sfi*I as recognised by pXJ1.1 and L1-4 might reflect methylation in recognition sequences of the type GGCmCGNNNNGGCC or GGCCNNNNGGCmCG. In this case observation of two bands due to an RFLP can be excluded, since two bands are also observed in an *Sfi*I digest of DNA from male blood cells (not shown). *Nae*I, although often producing bands on pulsed-field gels, gave a rather complex pattern with four and more bands (see Fig. 1*b*). These could not be interpreted without ambiguities. *Nae*I-fragments recognized by probes were therefore not considered in the construction of the map, and are not listed here.

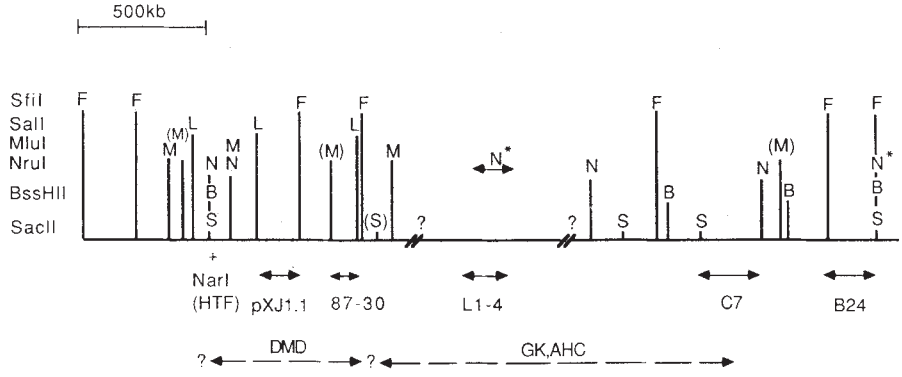


Fig. 2 Long range restriction map. The data for producing the map are summarized in Table 1. Weakly cut (highly methylated) sites are in brackets. The approximate locations of the DMD-gene(s) and of the glycerol kinase (GK) and congenital adrenal hypoplasia (AHC) genes are indicated below the map. Note that physical linkage of probes is only considered proven when at least three common bands generated by different enzymes were observed. Limits of localization for the probes used are indicated by arrows around the probe names below the map. In addition to the sizes of bands detected, the methylation status of a site was also considered for the construction of the map. The degree of methylation was estimated from the ratio of intensity of bands in double digests of enzymes cutting completely with the enzyme in question. Sizes were estimated with respect to yeast and λ size markers. It should be noted that the absolute sizes can only be considered to be approximate, because the amount of DNA loaded and pulse times[19] can slightly change the apparent size of a fragment. Comigration of two fragments was considered to be reliable, since the same blots were used for different probes. Since the cell line LCL127 is not homozygous for the X-chromosome, we can not rule out that some partial cutting is due to polymorphisms rather than methylation. Thus the map might have combined sites from two different chromosomes in a few instances. For the pERT87-30/pXJ1.1 cluster (Fig. 1*a*) subclones pERT87-30, localized about 50 kb distal of pERT87-15 (ref. 3 and Monaco and L. Kunkel, personal communication) and pXJ1.1 were used. Both pERT87-30 and pXJ1.1 recognize the same *Sal*I-fragment, but different *Sfi*I-fragments. Double digests of *Sfi*I and *Sal*I localize the *Sfi*I-site between the two probes approximately to the middle of the *Sal*I-fragment. Therefore, probe pXJ1.1 is located within 200 kb of the end of the large *Sfi*I fragment. Double digests with *Sfi*I/*Mlu*I and *Sfi*I/*Nru*I give the position of *Mlu*I and *Nru*I sites from the end of the *Sfi*I-fragment, similar to partial cosmid mapping[23]. A very weakly cutting (highly methylated) *Mlu*I-site within the *Sal*I fragment is detected in *Sal*I/*Mlu*I double digests. This *Mlu*I site cleaves the *Sal*I fragment into fragments of about 100 kb (hybridizing to pERT87-30) and of about 300 kb (hybridizing to pXJ1.1). The subclone pERT87-1, about 100 kb closer to pXJ1.1 than pERT87-30, recognizes the 300-kb *Sal*I-*Mlu*I fragment (not shown). This *Mlu*I-site is localized within the small *Sfi*I-fragment seen by pERT87-30, since it is also detected in a *Sfi*I/*Mlu*I digest by pERT87-30 as a weak *Sfi*I/*Mlu*I band of ~100 kb. Therefore the distance between pERT87-30 and pXJ1.1 is at least 100 kb, the distance between the weakly cut *Mlu*I site and the *Sfi*I-site separating the two probes. In the B24/C7 cluster (Fig. 1*b*), Probes B24 and C7 recognize the same *Sac*II fragments, but different *Sfi*I-fragments. A *Sac*II/*Sfi*I double digest then identifies a *Sac*II-site in the *Sfi*I-fragment of C7. The size of the common *Sac*II fragment predicts that the *Sfi*I fragments recognized by B24 and C7 are adjacent. There are different *Nru*I fragments recognized by these two probes. A *Nru*I/*Sfi*I double digest localizes one *Nru*I-site within the *Sfi*I-fragment seen by C7. Since the *Nru*I fragments are not shared, they point in opposite direction from this *Nru*I-site. This state of the map predicted a 250-kb *Nru*I/*Sac*II fragment recognized by C7, which could be confirmed (data not shown). Similarly, the *Bss*HII-sites were localized. Since *Nru*I does not cut the *Sfi*I-fragment seen by B24, the *Nru*I sites leading to a 450-kb *Nru*I fragment around B24 cannot be localized except for flanking the *Sfi*I fragment. A suggested *Nru*I-site (asterisk) is shown at the right hand side of the map. The additional *Nru*I-fragment of about 1,100 kb hybridizing to C7 seems not to be corecognized by B24, thus would be localized further to the left of the map (proximally). But as B24 hybridizes rather weakly, we can not completely rule out that the hybridization signal in this case was too weak to be detected. Therefore, the localization of the *Nru*I-site in the direction of L1-4 (asterisk) is considered not proven. The probe L1-4 could not link to any of the clusters (not shown). A 850-kb *Bss*HII fragment is recognized, and two *Sfi*I-fragments of 220 and 350 kb. As no bands are shared, the minimal distance between the *Mlu*I site distal of pERT87-30 and the *Nru*I site proximal to C7, is 500 kb, giving the distance of C7 to pERT87-30 as at least 1,000 kb.

recognized by L1-4 and any of the other probes. Taking the largest fragments not shared into account, the distance between the probes C7 and pERT87-30 must be at least 1,000 kb (Fig. 2).

When hybridizing pXJ1.1 to DNA restricted with *Sfi*I and several CpG-recognizing enzymes (*Nru*I, *Nar*I, *Sac*II and *Bss*HII) new fragments of identical size (390 kb) were detected in each case (see Fig. 1*c*). Since the probe pXJ1.1 is localized close to one end of the *Sfi*I fragment, these four sites must be localized, within the resolution of the gel, in close proximity to each other, 200-300 kb proximal to the probe pXJ1.1 (see map in Fig. 2). Such a clustering of sites would be expected in the case of an HTF (*Hpa*II tiny fragments) island[14,15], often found in close proximity to the 5′ end of genes. Very recently (data not shown) we found that at least some sites in the postulated HTF-island are cut to completion in DNA isolated from human sperm. The partial cutting of the sites observed here (Fig. 1*c*) is probably due to methylation of HTF-islands in inactivated X chromosomes[14]. Thus, with respect to methylation, this cluster behaves like known HTF-islands on the X-chromosome[14]. Since deletions in the pXJ1.1 region have a DMD phenotype as their only common effect[16,17], such an CpG-rich island could indicate the promoter sequence for the gene responsible for the DMD mutation. This would require the DMD gene to have a size of at least 600 kb (distance of the HTF island to the start point of the most distal published deletion[3]), with transcription proceeding in the direction centromere to telomere. Alternative models (for example, multiple clustered genes responsible for the same phenotype) can not be ruled out, but such a large gene size would be compatible with both the high incidence of DMD mutations and the observed distribution of deletions, translocations and other mutations associated with the disease throughout a very large chromosomal segment.

A complicating observation in the PFG analysis was partial cleavage of sites, resulting in multiple and weak bands recognized by these unique probes (compare, for example, the *Sfi*I-bands to all others in Fig. 1*a*, *b*). Partial restriction is caused by the sensitivity of all enzymes cutting rarely in the mammalian genome to CpG methylation in their recognition site. On the other hand, these partial digests added important information for mapping the region, since more sites can be analysed. Some of these sites might be due to polymorphisms since the lymphoblastoid cell line used contains four X chromosomes. This cell line was found to yield more visible bands than, for example male blood, as a result of a higher signal and/or hypomethylation. Nevertheless, major parts of the maps could be confirmed using DNA isolated from blood or sperm (not shown). All rare-cutting enzymes (except *Sfi*I, but see Fig. 1 legend) contain at least one, and usually two, CpG sequences in their recognition site. Reduced cleavage by CpG-recognizing enzymes is especially noticeable in the region surrounding the DMD locus, which appears to be depleted of unmethylated CpG-rich sites compared to other regions of the human genome which have been examined in our laboratory (M. Búcan *et al.*, unpublished data). For example, the enzyme *Not*I does not cleave within the entire region analysed. This obvious sparsity of unmethylated CpG-rich sequences might be correlated with a relatively low density of genes in this region, also indicated by the fact that large deletions around the DMD locus are associated with very few phenotypes[16,17]

We have described here partial physical maps covering ~3000 kb, about 2% of the human X chromosome. Such a map will be extremely useful in determining the exact size of the DMD gene(s), and will allow us to estimate the positions and extents of translocations and deletions in this area. The immediate positioning of new clones in the DMD area is now possible, with a resolution difficult to reach by purely genetic approaches.

We are especially grateful to Denise P. Barlow, who developed the PFG technology. We thank Anthony Monaco and Louis Kunkel, Peter Ray and Ron Worton, and Jean-Louis Mandel for providing the probes, Uta Francke for introducing us to the DMD problem, Marija Bucan for preparing some of the filters used, Gert-Jan van Ommen and Louis Kunkel for discussions, Denise P. Barlow, Michael Hortsch, Frank Michiels and Lisa Stubbs for critical reading of the manuscript, and Carol Stettner and other members of the electrical and mechanical workshops at the EMBL, for building the pulsed field gradient gel equipment.

Note added in proof: The recent publication of a cDNA from the DMD gene (A. M. Monaco *et al.*, *Nature* **323**, 646-650) confirms the direction of transcription and minimal gene size postulated here.

Received 28 July; accepted 29 October 1986.

1. Kedes, L. H. *Trends Genet.* **1**, 205-209 (1985).
2. Boyd, Y. & Buckle, V. J. *Clin. Genet.* **29**, 108-115 (1986).
3. Kunkel, L. M. *et al. Nature* **322**, 73-77 (1986).
4. Schwartz, D. C. & Cantor, C. R. *Cell* **37**, 67-75 (1984).
5. Smith, C. L. *et al.* in *Genetic Engineering* (eds J. K. Setlow & A. Hollander) Vol. 8, 45-70 (Plenum, New York, in the press).
6. Monaco, A. P. *et al. Nature* **316**, 842-845 (1985).
7. Ray, P. N. *et al. Nature* **318**, 672-675 (1985).
8. De Martinville, B. *et al. Am. J. hum. Genet.* **37**, 235-249 (1985).
9. Goodfellow, P. N., Davies, K. E., Ropers, H.-H. *HGM* **8**, *Cytogen. Cell genet.* **40**, 296-352 (1986).
10. Kunkel, L. M. *et al. Gene* **33**, 251-258 (1985).
11. van Ommen, G. J. B. *et al.* submitted.
12. Bertelson, K. *et al. J. med. Genet.* (in the press).
13. Dunger, D. B. *et al. Lancet* **i**, 585-587 (1986).
14. Bird, A. P. *Nature* **321**, 209-213 (1986).
15. Brown, W. A. R. & Bird, A. P. *Nature* **322**, 477-481 (1986).
16. Francke, U. *et al. Am. J. hum. Genet.* **37**, 250-267 (1985).
17. Wilcox, D. E. *et al. Hum. Genet.* **73**, 175-180 (1986).
18. Frischauf, A.-M. *et al. J. molec. Biol.* **170**, 827-842 (1983).
19. Bernards, A. *et al. Gene* **42**, 313-322 (1986).
20. Herrman, B. G. *et al. Cell* (in the press).
21. Feinberg, A. P. & Vogelstein, B., *Analyt. Biochem.* **132**, 6-13 (1983).
22. Herrmann, B. G. *et al. Cell* 469-476 (1986).
23. Rackwitz, H. R. *et al. Gene* **40**, 259-266 (1985).

Recombination between immunoglobulin variable region gene segments is enhanced by transcription

T. Keith Blackwell, Mark W. Moore*, George D. Yancopoulos, Heikyung Suh, Stuart Lutzker, Erik Selsing* & Frederick W. Alt

Department of Biochemistry and Microbiology, Institute for Cancer Research, Columbia University College of Physicians and Surgeons, 630 W. 168th Street, New York, New York 10032, USA
* Rosensteil Basic Medical Sciences Research Center and Department of Biology, Brandeis University, Waltham, Massachusetts 02254, USA

Immunoglobulin (Ig) variable (*V*) region genes are assembled in precursor B (pre-B) lymphocytes from multiple germline segments. The heavy-chain *V*-region gene is composed of variable (V_H), diversity (*D*) and joining (J_H) segments; kappa (κ) and lambda (λ) light-chain *V*-region genes have analogous V_L and J_L segments. Assembly of Ig *V*-gene segments, as well as those of the highly related T-cell receptor, is regulated at several levels and shows both stage and tissue specificity; for example Ig heavy-chain *V*-gene assembly precedes that of Ig light chains during B-cell differentiation[1]. Joining of all classes of *V*-gene segments involves conserved recognition sequences that are probably targets for a common recombinase[2]. Evidence has been presented suggesting that rearrangement of specific classes of segments is regulated by modulation of their accessibility to the recombinase[3]. To elucidate mechanisms which control *V*-region gene assembly, we have investigated the effect of flanking gene expression on the frequency at which introduced *V*-gene segments are assembled in pre-B cell lines. Our findings suggest that transcription may play a direct role in the regulation of immunoglobulin *V*-gene assembly.

Transformation by Abelson murine leukaemia virus (A-MuLV) generates pre-B cell lines; during culture some lines

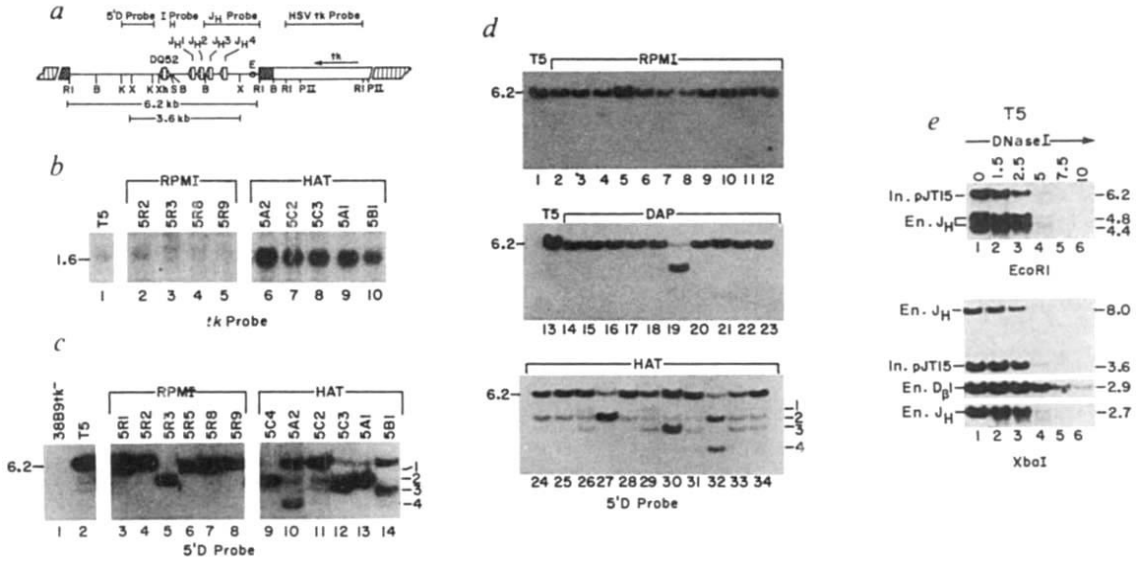


Fig. 1 Activation of transfected heavy-chain *V*-gene assembly. a, Map of the T5 integration. Line, murine DNA; checked box, plasmid DNA; open box, *tk* DNA; and lined box, cellular sequences. *D* and J_H segments with associated recognition sequences (triangles) are indicated. E, the heavy chain enhancer; arrow, *tk* gene. Restriction endonuclease sites are as follows: RI, *Eco*RI; B, *Bam*HI; K, *Kpn*I; X, *Xba*I; Xh, XhoI; S, *Sac*I; PII, *Pvu*II. *b*, *tk* expression in T5 derivatives. Total RNA (10 μg) from the cell lines indicated was assayed by Northern blotting for hybridization to a ^{32}P-labelled HSV *tk* probe as described[3]. *c*, Rearrangement of the integrated substrate. Genomic DNA (10 μg) from cell lines indicated was digested with *Eco*RI and analysed by Southern blotting for hybridization to a ^{32}P-labelled 5'D probe as described[4]. *DQ*52 to J_H 1-4 rearrangements are indicated by 1-4. *d*, Substrate rearrangements in T5 subclones isolated in RPMI (lanes 2-12), DAP (lanes 14-23) and HAT (lanes 23-34). *e*, DNase sensitivity of the integrated substrate. Nuclei isolated from T5 were incubated with 0-10 μg ml^{-1} of DNase I as described[2]. Genomic DNA (10 μg) was digested with either *Eco*RI or *Xba*I, hybridized to the J_H and D_B1 (ref. 2) probes, and assayed as in *c*. Endogenous (En.) and substrate (In.) fragments are indicated. Densitometry revealed that in lane 4 the pJT15 substrate and the 2.7-kb endogenous J_H-containing fragment retain 6 and 10%, respectively, of their lane 1 signal intensity, whereas the endogenous D_B1-containing fragment retains 81%. Fragment sizes are in kb.

Methods. The 38B9*tk*$^-$ cell line has been described[4]. T5 was derived by $CaPO_4$ co-transfection of 5×10^7 38B9 *tk*$^-$ cells with pJT15 and pSV2neo, followed by selection in G418 (ref. 4). HAT- or DAP-resistant clonal T5 derivatives were isolated in 24-well plates by plating 10^4-10^5 cells per well in 5×10^{-5} M DAP or in HAT[4]. For DNase sensitivity assays, T5 and 5R9 were expanded in RPMI and 5B1 in HAT.

assemble endogenous immunoglobulin *V*-region genes[1]. The Abelson line 38B9*tk*$^-$ has been shown to assemble endogenous and transfected heavy-chain *V*-gene segments[4]. To examine the effect of local gene expression on rearrangement activity, a substrate (pJT15) containing *DQ*52 and J_H segments plus the heavy-chain enhancer downstream from a herpes simplex virus thymidine kinase (*tk*) gene was introduced into 38B9*tk*$^-$ by unlinked co-transformation[4]. We used transfected clone T5, which had integrated a single copy of pJT15 (Fig. 1). As observed in 38B9 cells transfected with other *tk*-based constructs[4], *tk* messenger RNA was barely detectable in T5 cells (Fig. 1*b*, lane 1; see below). To isolate rare cells that expressed the *tk* gene, T5 populations were plated in HAT (hypoxanthine, aminopterin and thymidine) medium[4]; HAT-resistant clones occurred at a frequency of 10^{-4} to 10^{-5}, and all expressed normal 1.6 kb (kilobase) *tk* RNA from the *tk* promoter (Fig. 1*b*, lanes 6-10). Clones isolated in non-selective medium did not appear to express *tk* mRNA (Fig. 1*b*, lanes 2-5). The ability to isolate *tk*-expressing variants from transfected populations formed the basis of our experiments.

The region 5' of *DQ*52 (5'D probe, Fig. 1*a*) has been deleted from the 38B9 genome through endogenous joining[4] (Fig. 1*c*, lane 1). Thus, in T5 derivatives, the unrearranged pJT15 substrate was detected as a 5'D-positive, 6.2-kb *Eco*RI fragment (Fig. 1*c*, lane 2). In a given T5 cell, the *DQ*52 segment may join to any of the four J_H segments in the substrate; therefore, rearrangements which occurred during expansion of T5 subclones were detected as a mixture of four smaller, rearrangement-specific 5'D-positive *Eco*RI fragments (Fig. 1*c* and 1*d*). Five such rearranged fragments were molecularly cloned; sequence analysis indicated that all contained normal DJ_H joins (Fig. 2), some of which contained N regions[5]. Insertion of N regions into DJ_H rearrangements within the pJT15 construct correlated with terminal deoxynucleotidyl transferase RNA levels in clonal derivatives of a single line (Fig. 2), providing the strongest evidence to date linking this enzyme with N-region addition[2,5].

Genomic DNA from the initial T5 population contained a very low but detectable level of specific rearrangements (Fig.

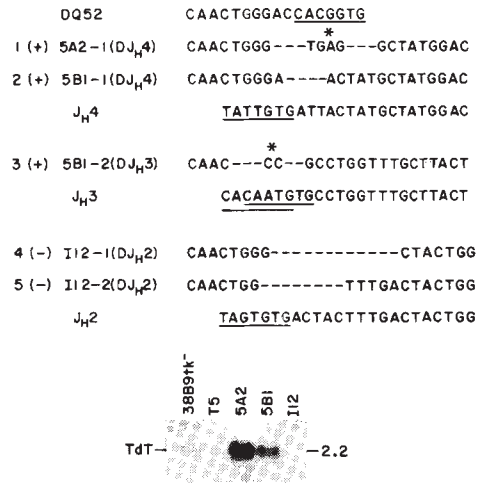


Fig. 2 Nucleotide sequences of *DQ*52 to J_H joins. *Eco*RI fragments hybridizing to the 5'D probe but not the I probe (Fig. 1*a*) were isolated from Charon 16A libraries of 5A2, 5B1 and I12 DNA as described[4]. Joins isolated from terminal deoxyribosyl transferase (TdT)-positive (+) and TdT-negative (−) lines are noted; nucleotide sequences were determined as described[4]. Recognition heptamers are underlined, and joins aligned with germline sequences so that bases lost during joining are apparent. Asterisks, N regions. Lower right panel: Total RNA from the lines indicated was assayed for hybridization to a TdT cDNA clone as described[2]. The TdT transcript and its approximate size in kb are indicated.

Table 1 Substrate rearrangement correlation with HAT resistance

	Experiment no. 1		Experiment No. 2		Experiment No. 3		Experiment No. 4		Total	
	RPMI	HAT	RPMI	HAT	RPMI	HAT	RPMI	HAT	RPMI	HAT
Subclones examined	7	5	6	9	11	11	11	11	35	36
Unrearranged	4	0	5	0	10	2	10	0	29	2
Single rearrangement	2	0	1	0	1	1	0	0	4	1
Rearranged*	1	5	0	9	0	8	1	11	2	33

*DQ*52 to J_H joining was assayed in clonal T5 derivatives in four separate experiments.
* A level of *DQ*52 to J_H joining within the population at or greater than that of Fig. 1*d*, lane 4.

1*c*, lane 2), suggesting the presence of a minor subpopulation that had joined introduced *D* and J_H segments. Clonal populations derived in RPMI did not express the *tk* gene or show evidence of substrate rearrangement (Fig. 1*c* and *d*); a few tk^- clones had a single *D* to J_H join (Fig. 1*c*, lane 5) demonstrating that *D* to J_H rearrangement *per se* did not initiate expression of the contiguous *tk* gene (Fig. 1*b*, lane 3). Such joins probably occurred before subcloning, as cells containing the 6.2-kb unrearranged substrate were absent. In contrast, *tk*-expressing clonal populations isolated in HAT medium displayed a striking level of rearrangement within the substrate (Fig. 1*c* and *d*). Most HAT-resistant clonal populations also contained many cells with an unrearranged 6.2-kb substrate, indicating each cell derived from a progenitor that expressed an unrearranged *tk* substrate, and that joining occurred during expansion of the population (Fig. 1*c* and *d*). Subcloning of initial HAT-resistant clones confirmed that *D* to J_H joining within the substrate occurred frequently in *tk*-expressing derivatives (not shown). HAT resistance (*tk* expression) was correlated with rearrangement of the introduced *D* and J_H segments in four separate experiments (Table 1).

To investigate whether increased recombination frequencies in HAT-resistant cells were associated with *tk* expression rather than other potential effects of selection (such as interference with nucleotide metabolism), T5 cells were also grown in 2,6-diaminopurine (DAP) which selects against cells expressing adenosine phosphoribosyl transferase[6]; conditions were employed such that survival frequency in DAP was less than in HAT. Surviving clones (Fig. 1*d*, lanes 14–23) showed low levels of rearrangement, as did those derived in RPMI (Fig. 1*d*, lanes 2–12). Therefore, the increased frequency of *DQ*52 to J_H rearrangement in HAT-resistant clones is associated specifically with expression of the flanking *tk* gene.

The 'active' chromatin state required for expression of eukaryotic genes is reflected by its sensitivity in isolated nuclei to digestion with DNase I (DNase sensitivity)[7,8]. To define the requirements for *V*-gene assembly further, we examined relative DNase sensitivity of the pJT15 construct in the nuclei of the T5 line, the HAT-derived 5B1 line (which expresses the *tk* gene and actively rearranges the substrate) and the RPMI-derived 5R9 line (which does not express the *tk* gene or rearrange the substrate). In all three the DNase sensitivity was similar to that of the active endogenous J_H loci; all were DNase sensitive relative to the inactive[2] D_B1 locus of the T-cell receptor (Fig. 1*e*, lanes 1–6; only T5 data are shown). Thus, DNase sensitivity *per se* is not responsible for active rearrangement of the introduced construct.

To analyse further the basis for low frequency expression of introduced HSV *tk* genes in 38B9 and the relationship between increased *tk* expression and recombination in HAT-resistant derivatives, we analysed a 38B9 *tk*-transformant, L1, which contained multiple intact integrations of a construct containing $V_\lambda 1$ and $J_\lambda 1$ elements flanking a *tk* gene (pLT2, Fig. 3*a*). In this case, transformation was accomplished by unlinked co-electroporation[2]. Assuming that low frequency expression of a co-transfected *tk* gene in 38B9tk^- derivatives resulted from a '*cis*-acting' transcriptional modulation, it would be expected that HAT selection would isolate a population of cells which expressed the *tk* gene in only one of the integrated substrate copies. Because expressed and non-expressed construct copies can be distinguished (see below), we could directly compare the effect of differential expression on rearrangement activity of identical constructs within single cells. Although the 38B9 line is in the heavy-chain phase of its *V*-region gene assembly programme and does not rearrange endogenous λ light chain genes at a detectable level, earlier evidence suggests that all immunoglobulin gene segments are rearranged by a common recombinase[2]; therefore, introduced λ gene segments, if accessible, should be rearranged in this line. Integrated pLT2 copies were detected in genomic DNA of L1 derivatives as $V_\lambda 1$-positive *Pvu*II fragments of different sizes; a *V–J* rearrangement (for example VJ1, Fig. 3*c*) was detected as a $V_\lambda 1$-, $J_\lambda 1$-positive *Pvu*II fragment 1.7 kb smaller than the $V_\lambda 1$-positive *Pvu*II fragment containing the copy from which it was derived (for example RC1, Fig. 3*c*). Nucleotide sequence analysis of a *VJ*1 join molecularly cloned from L1HB confirmed that it occurred by the normal mechanism (Fig. 3*d*).

A population of L1 cells that expressed the introduced *tk* gene was isolated in HAT (L1H, Fig. 3*b*); these HAT-resistant cells again appeared at a frequency of 10^{-4} to 10^{-5}. With respect to integrated construct copies, the genomic DNA of the L1H population appeared identical to that of L1 except for complete loss of a substrate copy which was identified by a 4.5 kb $V_\lambda 1$-positive *Pvu*II fragment (see Fig. 3*c*). Assuming a common recombinase[2], deletion of the expressed *tk* gene(s) as a result of V_λ to J_λ joining within the construct would be expected to be the major mechanism of BUdR(bromodeoxyuridine)-resistance in these lines[4]. Removal of L1H from HAT (L1HT, Fig. 3*b*) followed by selection in BUdR yielded the tk^- population L1HB. These cells had completely rearranged a single integrated construct (RC1) to its product *VJ* join (*VJ*1) (Fig. 3*c*, lanes 3 and 7), whereas other full length copies remained intact. This identifies RC1 as the *tk* copy expressed in this subclone and confirms that the *tk* mRNA found in HAT-resistant subclones derives from a single integrated substrate. This rules out potential post-transcriptional (*trans*-acting) mechanisms for increased *tk* expression that would affect expression from all copies, and strongly suggests that low frequency expression of transfected *tk* genes in these lines is mediated at the transcriptional level.

To identify independently pLT2 construct loci that are expressed, we measured relative DNase sensitivity of the integrated constructs. L1H was expanded in non-selective medium, generating population L1HT2 (Fig. 3*b*), in which RC1 undergoes rearrangement (Fig. 3*e*, lane 1; see below). In L1HT2, both RC1 and its rearranged product *VJ*1 are clearly DNase sensitive relative to endogenous genes and non-rearranging pLT2 integrations (Fig. 3*e*, lanes 1–6). This provides additional evidence that *tk* expression occurs primarily from RC1 in L1H and that activation of this copy is at the transcriptional level. Significantly, after a brief passage in non-selective medium, L1H cells converted the RC1 integration to its derivative rearrangement; however, no rearrangement of other integrated copies was observed (Fig. 3*c*, lanes 4 and 8). Thus, the transcribed construct copy has a high rate of rearrangement but non-expressed copies in the same line do not; this is not because of defects in their

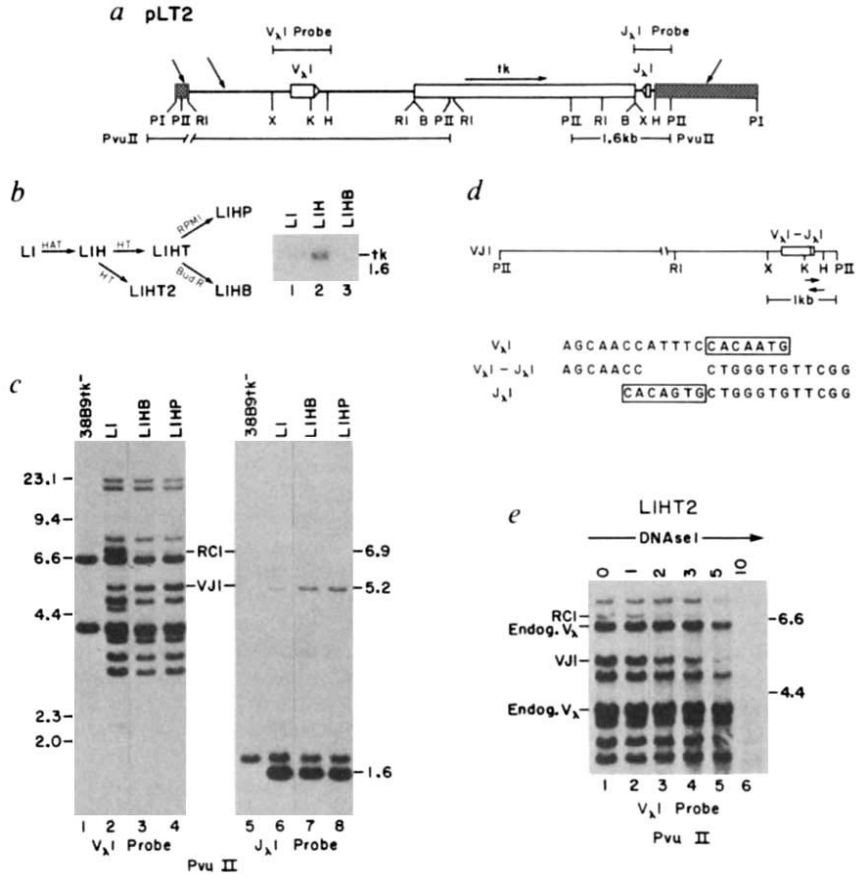


Fig. 3 Rearrangement of transcribed light chain substrates. *a*, Map of the pLT2 substrate. Structures and restriction endonuclease sites as in Fig. 1*a* (also H, *Hin*dIII; PI, *Pvu*I). Arrows above map, regions within which pLT2 copies integrated into cellular DNA. *b*, Selection of L1 derivatives and *tk* expression. Total RNA (10 μg) was assayed for *tk* expression as in Fig. 1*b*. *c*, Rearrangement of the pLT2 substrate. Genomic DNA (10 μg) was digested with *Pvu*II and analysed as in Fig. 1*c* for hybridization to the probes indicated. Sizes are in kb. *d*, Restriction map and nucleotide sequence of a *VJ*1 rearrangement. To isolate this join, L1HB genomic DNA was digested with *Pvu*II, *Eco*RI linkers added and fragments cloned into Charon 16A and screened with $V_\lambda 1$ and $J_\lambda 1$ probes. The join is aligned with germline sequences so that the bases lost are apparent. *e*, DNase sensitivity of integrated and endogenous (Endog.) λ loci. Nuclei from L1HT2 were digested with 0–10 μg ml^{-1} of DNase I; genomic DNA was then digested with *Pvu*II and assayed for hybridization to the $V_\lambda 1$ probe as described in Fig. 1*c*. Fragments corresponding to RC1, *VJ*1 and the endogenous V_λ loci are indicated. Densitometry revealed that in lane 4, RC1 and *VJ*1 retained 17 and 36%, respectively, of their lane 1 signal intensity. In lane 4 the pLT2 copy which is slightly larger than RC1, and the endogenous 6.4-kb V_λ-containing fragment retained 87 and 97%, respectively, of their original intensities. Sizes are in kb.

Methods. pLT2 was derived from pUC13. pLTRneo contains the TN*5neo* resistance gene transcribed from a Moloney-MuLV LTR promoter. L1 was derived by electroporation[2] of 7.5×10^6 38B9 *tk*$^-$ cells with 10 μg of pLTRneo and 40 μg of pLT2 which had been linearized with *Pvu*I. Transformants were isolated in G418 as in Fig. 1*a*. L1 was passaged through HAT, HT and 10^{-4} M BUdR as described[4]. Other analyses were as described in Figs 1 and 2.

structure, as some different copies were expressed and rearranged in independent L1 derivatives (not shown).

Identical *tk*-based recombination substrates are expressed frequently (>25%) when introduced into myeloma (J558) or L cell lines by the same procedures (not shown) that lead to infrequent expression (10^{-4}–10^{-5}) in pre-B lines. This may be related to the relatively low transformation frequency of pre-B lines[4]. Current results suggest that the block in pre-B cells is transcriptional but does not result from lack of general DNase sensitivity (Fig. 1) or obvious methylation-related phenomena (not shown). Significantly, we found no site-specific recombination events in L cell or J558 transformants when they were subjected to analyses identical to those described above (data not shown), suggesting that recombinase activity may be regulated with respect to stage and tissue specificity. The assay would have detected recombinase activity 3–4 orders of magnitude lower than found in several pre-B cell lines. However, many more different lines must be assayed to verify this point.

In T5 derivatives, the pJT15 substrate is of equivalent DNase-sensitivity to the endogenous J_H loci regardless of *tk* gene expression (Fig. 1*e*); this sensitivity is probably conferred by the associated heavy-chain enhancer[9–10] which is active in pre-B cells[11]. Therefore, activity of the enhancer element (and DNase-sensitive structure *per se*) on the pJT15 substrate appears insufficient to confer direct activation of *D* to J_H rearrangement. T-cell receptor *V*-gene segments introduced into 38B9 within an *expressed tk*-based recombination substrate were DNase sensitive and had a high spontaneous rearrangement rate[2]. In the λ substrate which did not contain known enhancer elements, DNase sensitivity (and rearrangement) appeared to occur only in association with transcription of the *tk* gene. Apparently, efficient assembly of *V* segments requires either transcription or factors closely associated with transcription. Our results do not preclude a regulatory role for enhancer elements with respect to rearrangement; they may allow transcription or closely-associated events which then promote rearrangement. The lack of a clear function for known or potential products of germline immunoglobulin or T-cell receptor gene transcripts supports a direct role for transcription[1]. A relationship between local gene expression and recombination has been observed in the context of other genomic rearrangement events, such as the yeast mating-type locus switch[12] and the immunoglobulin class switch [13,14]. Transcription could influence accessibility for rearrangement in several ways. The recombination mechanism might involve transcription itself; for example DNA unwinding by RNA polymerase[15] could be required for initiation of recombination. Alternatively, transcription of these loci might be associated with structural or nuclear locational changes in DNA that are not detected by general DNase sensitivity and not solely mediated by the enhancer[8,16].

We thank Y. Rosenberg, B. Malynn and P. Ferrier for critically reading this manuscript, and C. S. H. Young, S. Morrison and S. Silverstein for helpful discussions. Supported by NIH Grants AI-20047, CA-40427 and CA-2112-09, and a Searle Scholars Award to F.W.A. (who is an Irma T. Hirschl Career Scientist and a Malinckrodt Scholar) and by NIH grant CA-36370 and NSF grant DCB-8302110 to E.W.S. T.K.B., S.L. and G.D.Y. were supported by NIH training grant GM-07367, and M.W.M. by NIH training grant AI-07069.

Received 26 June; accepted 23 September 1986.

1. Alt, F. W., Blackwell, T. K., De Pinho, R.A., Reth, M. G. & Yancopoulos, G. D. *Immun. Rev.* **89**, 5–30 (1986).
2. Yancopoulos, G. D., Blackwell, T. K., Suh, H., Hood, L. & Alt, F. W. *Cell* **44**, 251–259 (1986).
3. Yancopoulos, G. D. & Alt, F. W. *Cell* **40**, 271–281 (1985).
4. Blackwell, T. K. & Alt, F. W. *Cell* **37**, 105–112 (1984).
5. Alt, F. W. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4118–4122 (1982).

6. Jones, G. E. & Sargent, P. A. *Cell* **2**, 43-54 (1974).
7. Weintraub, H. & Groudine, M. *Science* **193**, 848-856 (1976).
8. Weintraub, H. *Cell* **42**, 705-711 (1986).
9. Banerji, J., Olson, L. & Schaffner, W. *Cell* **33**, 729-740 (1983).
10. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. *Cell* **33**, 717-728 (1983).
11. Gerster, T., Picard, D. & Schaffner, W. *Cell* **45**, 45-32 (1986).
12. Klar, A. J. S., Strathern, J. N. & Abraham, J. A. *Cold Spring Harb. Symp. quant. Biol.* **49**, 77-88 (1984).
13. Stavnezer-Nordgren, J. & Sirlin, S. *EMBO J.* **5**, 95-102 (1986).
14. Yancopoulos, *et al. EMBO J.* (in the press).
15. Fisher, M. L. *Nature* **299**, 105-106 (1982).
16. Hutchinson, N. & Weintraub, H. *Cell* **43**, 471-482 (1985).

Effect of mutations at the lariat branch acceptor site on β-globin pre-mRNA splicing *in vitro*

H. Hornig, M. Aebi & C. Weissmann

Institut für Molekularbiologie I, Universität Zürich, 8093 Zürich, Switzerland

Introns are excised from full-length transcripts (pre-messenger RNAs) of eukaryotic genes in two steps[1,2]. First, the pre-mRNA is cleaved at the 5′ splice site and a branched (lariat) intermediate is formed[1-6]. Then, cleavage at the 3′ splice site and ligation of the two exons leads to the release of the lariat intron. The intron sequence which accepts the 5′ end to form the lariat branch is strictly conserved in yeast[7-9], but shows more variation in eukaryotes[10,11]. To investigate the requirements for branch formation in eukaryotes further, we have studied *in vitro* splicing of a rabbit globin gene intron with mutations of the normal branch-accepting adenosine nucleotide. We conclude that all four nucleotides can serve as branch acceptors, but that A and C are preferred to G and U in lariat formation. Mutation of the normal A to G or U can lead to an A residue one nucleotide upstream of the normal branch site being used instead. Only branches to A or C participate efficiently in the second splicing step.

Lariat formation[1-5] results from the 2-5′ linkage of the 5′ end of the intron to an adenosine residue close to the 3′ splice site[6]. In the case of yeast nuclear pre-mRNA splicing, efficient lariat formation depends stringently on the strictly conserved TACTAAC-box, where the underlined 3′-proximal A residue serves as branch acceptor[4,5]. Deletions or point mutations within this sequence abolish splicing[7-9]. In contrast, the branch acceptor sequence of higher eukaryotes is poorly conserved[10,11] and deletions of this sequence lead to the activation of cryptic branch points which allow normal splicing, albeit at a reduced rate[12,13]. There seems to be a distance constraint as regards branch formation, in that the acceptor sequences identified so far lie within 18-37 nucleotides upstream of the 3′ splice site[11,12].

The branch acceptor nucleotide of the large rabbit β-globin intron has been determined *in vivo*[3] as well as *in vitro* [13] to be predominantly the underlined A residue and to a lesser extent the A residue preceding it, in the sequence GCTAAC. In order to study the nucleotide requirement at the branch point, we have converted the principal branch acceptor A residue (position 1,036)[14] to either G, C or T by site-directed mutagenesis. The mutated β-globin sequences (*Bam*HI-*Sma*I fragments, Fig. 1*A*) were substituted for their wild-type counterparts in plasmid SP64/Rchrβ G[13], which allows *in vitro* synthesis of a globin pre-mRNA containing both the small and the large intron (Fig. 1*A*). Wild-type, as well as the three mutant pre-mRNAs 1,036 A→G, 1,036 A→C and 1,036 A→T, yielded the correctly-spliced product of 536-nucleotide length when spliced *in vitro* (Fig. 1*B*). The rate of splicing of the large introns of the mutants, as determined within the linear period of the reaction (60-120 min), was about 50% that of wild type (Table 1). In the case of the mutant substrates 1,036 A→G and A→T a slight accumulation of the lariat intermediate of the large intron relative to the lariat product was observed (Fig. 1*B*).

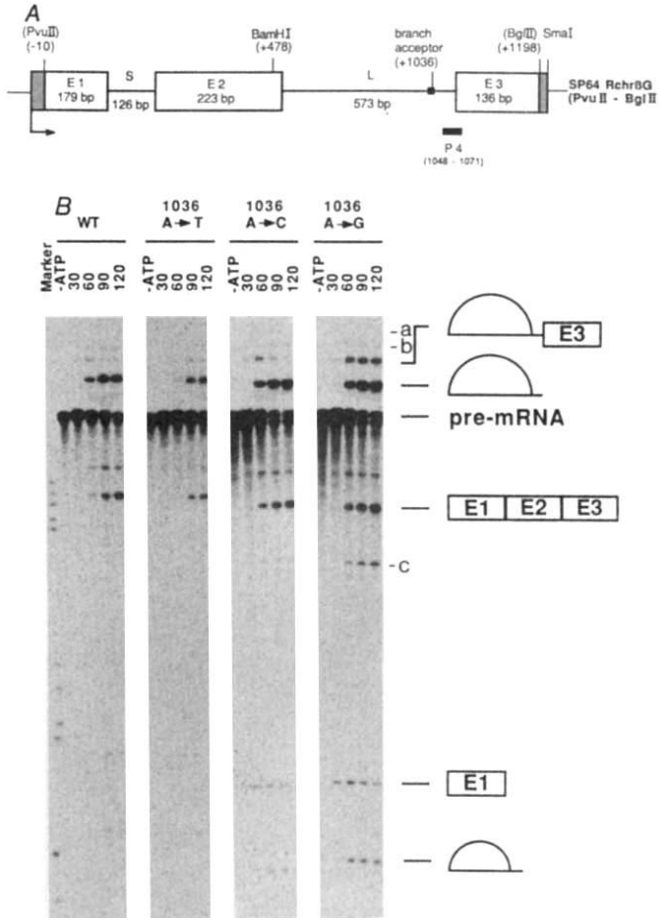


Fig. 1 Analysis of the *in vitro* splicing products of pre-mRNAs with branch acceptor mutations. *A*, Partial map of the *Pvu*II-*Bgl*IIRchrβG DNA fragment[14] cloned into pSP64 for synthesis of pre-mRNA *in vitro*. The construction of the wild-type plasmid has been described[13]. Numbering is relative to the natural cap site (+1). Initiation of transcription is indicated by an arrow. Point mutations at position 1,036 were obtained by oligonucleotide-directed mutagenesis[19]. The mutated *Bam*HI-*Sma*I fragments were substituted for their wild-type counterparts and the resulting plasmids used as template for *in vitro* synthesis of pre-mRNA[22,23]. The resulting transcripts contained 24 and 3 nucleotides of polylinker sequences at the 5′ and 3′ ends, respectively. The location of the oligonucleotide primer used in primer elongation is shown. The drawing is not to scale. *B*, *In vitro* splicing of wild-type and branch-point-mutant pre-mRNA. The ^{32}P-UMP-labelled pre-mRNA (specific activity 2×10^6 c.p.m. pmol^{-1} RNA) of the wild type and the mutants indicated above each panel were incubated with HeLa-cell nuclear extract[6,24] for the time (min.) indicated, and analysed on a 3.2% denaturing polyacrylamide gel[25]. The relevant products and intermediates are indicated; for a description of structures a (superlariat intermediate), b (superlariat) and c (E1-E2), see refs 13 and 19. '-ATP', ATP and creatine phosphate omitted and the RNA incubated with HeLa-cell nuclear extract for 120 min. Marker, 5′-^{32}P-labelled *Bsp*I-digested pBR322.

As shown previously[13], the exact position of the branch point can be determined by primer extension analysis using the 24-nucleotide *primer 4* complementary to the region 1,048-1,071 (ref. 14), which is 12 nucleotides downstream of the branch acceptor A nucleotide (Fig. 1*A*). Reverse transcription of branched RNA molecules, either the excised lariat intron or the lariat intermediate containing the third exon, is arrested at the position of the branch[2]. Wild-type RNA yields a major primer extension product, indicative for the branch at position 1,036 (ref. 13) and a minor product extending to position 1,035, the

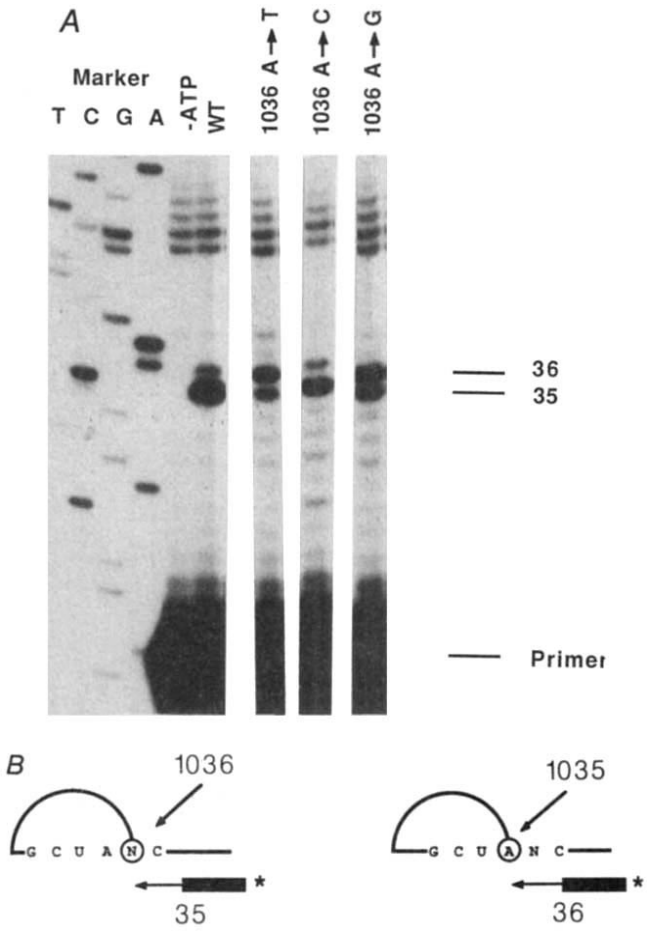


Fig. 2 Analysis of branch point use by primer extension. *A*, Wild-type and mutant pre-mRNA (0.1 pmol each) were incubated with HeLa-cell nuclear extract and RNA was extracted after 1.5 h. For annealing of the 5′-^{32}P-labelled 24-nucleotide *primer 4* (specific activity 4×10^6 c.p.m. pmol^{-1}) and primer extension using reverse transcriptase, see ref. 13. Extension products were analysed on a 12% denaturing polyacrylamide gel[25]. Marker, dideoxy chain termination sequence[26] of single-stranded M13 mp8 RchrβG (*Bam*HI-*Eco*RI) using the universal N15 primer[27]. *B*, Scheme of the expected primer extension products indicative for branch-point formation at position 1,036 and 1,035. Black box, 5′-^{32}P-labelled primer (*, position of label).

A residue immediately 5′ to the major branch point (Fig. 2*A*, *B*), in a 10:1 ratio (Table 1). In the case of the mutant pre-mRNAs 1,036 A→T and A→G, about 60% of the branches were displaced to the A residue at position 1,035, while the remaining 40% were connected to the mutated nucleotide at position 1,036 (Fig. 2*A*, Table 1). The 1,036 A→C mutant yielded predominantly C-branched molecules (85%), while the neighbouring upstream A-nucleotide was used in 15% of the lariat molecules. We conclude that all four nucleotides can serve as branch acceptors, albeit with different efficiencies, namely A > C > G, U.

Since in the above experiment a mixture of lariat intermediate and lariat product had been analysed, it was not clear whether the molecules that were branched at the mutated nucleotide were capable of undergoing cleavage and exon joining at the 3′ splice site, or whether they accumulated as dead end lariat intermediates. We isolated the lariat product as well as the lariat intermediate by polyacrylamide gel electrophoresis. In the case of wild type, lariat intermediate and lariat product yielded the same pattern of primer extension products (Fig. 3): the A residue at position 1,036 was the major branch acceptor in both cases. Different results were obtained with the two mutants 1,036 A→G

Table 1 Relative rate of splicing and use of different branch points

Nucleotide at position 1,036 (N)	Relative rate of splicing	Relative use of branch point: 1,036 GCUA(N)C—	Relative use of branch point: 1,035 GCU(A)NC—
A(wt)	100%	90%	10%
T	60%	35%	65%
C	45%	85%	15%
G	55%	40%	60%

The nucleotide N (in DNA) was substituted with the nucleotides listed in column 1. Appropriate bands were cut out of a polyacrylamide gel like the one in Fig. 1*B* and counted in a liquid scintillation counter. The relative splicing rate was calculated from the ratio of fully spliced product E1-E2-E3 to small lariat intron at different times. The rate of excision of the first intron is independent of the rate of excision of the second intron, as determined by quantifying the data in Fig. 3 of ref. 19, and can therefore serve as a reference value. This ratio is taken as 100% for the wild type. To calculate relative branch point use, the two primer extension products (Fig. 2*A*) indicative for the two branch points were excised from the gel and the radioactivity was determined. wt, wild-type nucleotide.

and 1,036 A→T: whereas the lariat products were branched almost exclusively at the A residue (position 1,035) immediately upstream of the mutated position, the lariat intermediate was branched predominantly at the mutated position itself. In contrast, the lariat intermediate and lariat intron derived from the mutant pre-mRNA 1,036 A→C yielded mainly lariat product branched at the mutated residue (Table 2). This analysis assumes that the lariat products of all the pre-mRNAs examined are equally stable in the *in vitro* system and are therefore representative of the processed lariat intermediates. We believe this to be the case because the ratio of lariat product to mature mRNA was about the same for the wild type and all the mutants throughout the reaction (Fig. 1*B*). We have also determined that the deproteinized mutant lariat products are more resistant to debranching by nuclear extract than the wild-type lariat (the relative rates decrease in the order wild type > 1,036 A→T, 1,036 A→G > 1,036 A→C), however, as native lariats are resistant to debranching[15], this finding is not necessarily relevant to the question of selective degradation.

We conclude from these results that lariat intermediate branched at a C-nucleotide is functional in the second step of the splicing reaction, whereas G- or U-branches are strongly inhibitory.

This observation is in agreement with the results of Wallace and Edmonds[16], who reported that the predominant branched nucleotide in nuclear RNA was an A residue. In contrast to previous reports, more detailed analysis revealed that branched C-residues were the next most frequently found (J. C. Wallace and M. Edmonds, personal communication). Moreover, a naturally-occurring C-branched lariat of the first intron of the human growth hormone pre-mRNA has been identified (K. Hartmuth and A. Barta, personal communication).

An A→G transition of the branch acceptor of the small intron of the human β-globin gene has been shown to lead to a fivefold reduction of the splicing efficiency *in vitro*[12]. The mutated branch acceptor nucleotide (37 nucleotides upstream of the 3′ splice site) was not used in lariat formation. Instead, a cryptic branch acceptor sequence 24 nucleotides upstream of the 3′ splice site was activated. The difference between these results and ours may be due to the fact that different introns were examined in the two studies. The affinity of the original branch acceptor sequence for the human β-globin small intron to the component recognizing it may be weaker than that of the rabbit sequence we have studied. Further weakening of this affinity by an A→G transition of the branchpoint might lead to the use of a cryptic site which has a more favourable sequence and/or is closer to the 3′ splice site. There seems to be no stringent sequence

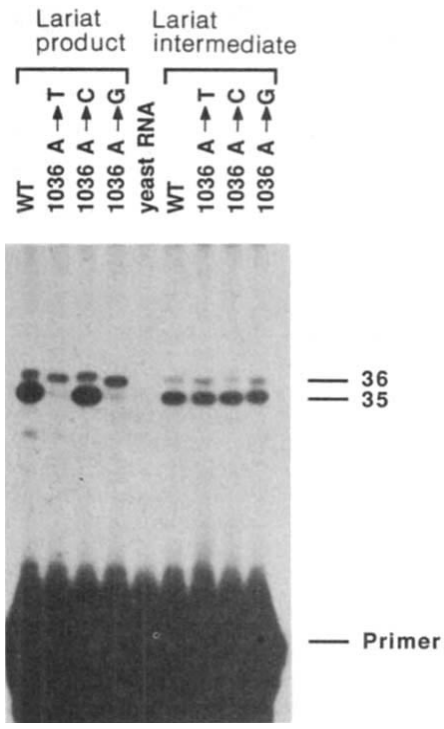


Fig. 3 Branching in lariat product and lariat intermediate derived from pre-mRNAs with branch point mutations, analysed by primer extension. Wild-type and mutant pre-mRNAs (4 pmol each) were spliced *in vitro* as described in Fig. 1. Lariat product and lariat intermediate were isolated by preparative polyacrylamide gel electrophoresis and analysed by primer extension using the 5′-^{32}P-labelled *primer 4* (see Fig. 1*A*). The 35- and 36-nucleotide extension products indicate branch points at positions 1,036 and 1,035, respectively (see Fig. 2*B*). Yeast RNA was used as carrier RNA and served as a negative control.

requirement for branch acceptor sequences, as indicated by the variety of natural and cryptic branch acceptor sequences reported to date[12,13].

All mutations of the branch acceptor A of the rabbit β-globin large intron led to a 50% reduction in the rate of splicing as compared to wild type. Since complex formation around the branch acceptor sequence is an early step in the splicing reaction[17,18], the observed rate in the three mutant RNA splicing reactions may be caused by a lower affinity of the splicing components for the branch acceptor sequence, or by a decreased rate of branch formation. Our results do not allow us to differentiate between these two possibilities. As the branch is nevertheless formed at the mutated position or at the A-nucleotide just upstream of the normal branch site, and no activation of cryptic branch acceptor sequences is detected, we conclude that the mutated sequence is still functional, which confirms that there is no stringent sequence requirement for branch formation. Although an A- or C-residue is favoured over G or U, branch formation is possible at all four nucleotides. In the case of the mutants, preference for an A residue may cause the activation of the A-nucleotide at position 1,035, one nucleotide upstream of the normal branch site, as branch acceptor.

It is surprising that the sequence requirements for the first step of the splicing reaction, where the branch region is intimately involved, are lower than for the second step, where it is at a distance from the reacting nucleotides. In the first step, recognition of the branch-forming region by splicing components may not be stringent, and /or the reactivity of the branch-forming nucleotide as regards its capacity of attacking the phosphodiester bond upstream of the 5′ terminal G residue of the intron may not be very sequence specific. It is even possible

Table 2 Different frequencies of branchpoints in lariat products and lariat intermediates

Nucleotide at position 1,036 (N)	Relative use of branch point: Lariat product, 1,036 (GCUA(N)C)	Lariat product, 1,035 (GCU(A)NC)	Lariat intermediate, 1,036 (GCUA(N)C)	Lariat intermediate, 1,035 (GCU(A)NC)
A(wt)	80%	20%	80%	20%
T	10%	90%	80%	20%
C	80%	20%	90%	10%
G	20%	80%	80%	20%

The primer extension analysis of Fig. 3 was quantified by cutting the cDNA corresponding to different branch points out of the gel, and determining the radioactivity in a scintillation counter. wt, wild-type nucleotide.

that assembly of the nucleoprotein complex at the branch-forming region takes place at a fixed distance from the polypyrimidine stretch at the end of the intron, regardless of the actual sequence at that position, and that branch formation occurs from the closest suitable nucleotide.

As regards the second step of splicing, it is not immediately clear why 3′ cleavage and exon joining should be affected by a mutation at the branch point, which is some 20–30 nucleotides upstream of the 3′ cleavage site. It is important to note that certain mutations at the 5′ end of the β-globin large intron, namely those converting the 5′ GT to AT, GA and GG, also prevent the second step of splicing, while allowing the first[19]. A similar observation has been made for a yeast 5′ GT→AT mutant[20]. In yeast, moreover, the requirements at the branch point (TACTAAC) are apparently even stricter than in higher eukaryotes, because an A→C substitution of the branch acceptor nucleotide not only slows down branch formation to the C-residue but very strongly reduces the second step of splicing[21]. It would thus seem that at least three nucleotides close to the branch, namely the branch point nucleotide and the two closest residues linked to it via the 2′–5′ linkage, are required for the second splicing step. This requirement could come about either because a component that binds to the branch region in connection with the first splicing step does not undergo an essential conformational change, or fails to dissociate or translocate or because a component required for the second step of the splicing reaction is unable to recognize the mutated sequence.

Supported by grants from the Schweizerische Nationalfonds and the Kanton of Zürich. H. H. was a recipient of an EMBO Long-Term Fellowship. We thank D. Schuppli for technical assistance and F. Ochsenbein for preparing the photographs.

Received 27 June; accepted 14 October 1986.

1. Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F. & Sharp, P. A. *Science* **225**, 898–903 (1984).
2. Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. *Cell* **38**, 317–331 (1984).
3. Zeitlin, S. & Efstratiadis, A. *Cell* **39**, 589–602 (1984).
4. Domdey, H. *et al. Cell* **39**, 611–621 (1984).
5. Rodriguez, J. R., Pikielny, C. W. & Rosbash, M. *Cell* **39**, 603–610 (1984).
6. Konarska, M. M., Grabowski, P. J., Padgett, R. A. & Sharp, P. A. *Nature* **313**, 552–557 (1985).
7. Langford, C. J. & Gallwitz, D. *Cell* **33**, 519–527 (1983).
8. Pikielny, C. W., Teem, J. L. & Rosbash, M. *Cell* **34**, 395–403 (1983).
9. Langford, C. J., Klinz, F.-J., Donath, C. & Gallwitz, D. *Cell* **36**, 645–653 (1984).
10. Keller, E. B. & Noon, W. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7417–7420 (1984).
11. Reed, R. & Maniatis, T. *Cell* **41**, 119–126 (1985).
12. Ruskin, B., Greene, J. M. & Green, M. R. *Cell* **41**, 833–844 (1985).
13. Padgett, R. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 8349–8353 (1985).
14. Van Ooyen, A., van den Berg, J., Mantei, N. & Weissmann, C. *Science* **206**, 337–344 (1979).
15. Ruskin, B. & Green, M. R. *Science* **229**, 135–140 (1985).
16. Wallace, J. C. & Edmonds, M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 950–954 (1983).
17. Black, D. L., Chabot, B. & Steitz, J. A. *Cell* **42**, 737–750 (1985).
18. Ruskin, B. & Green, M. R. *Cell* **43**, 131–142 (1985).
19. Aebi, M., Hornig, H., Padgett, R. A., Reiser, J. & Weissmann, C. *Cell* (in the press).
20. Newman, A. J., Lin, R. J., Cheng, S. C. & Abelson, J. *Cell* **42**, 335–344 (1985).
21. Jacquier, A. & Rosbash, A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5835–5839 (1986).
22. Melton, D. A. *et al. Nucleic Acids Res.* **9**, 7035–7056 (1984).
23. Konarska, M. M., Padgett, R. A. & Sharp, P. A. *Cell* **38**, 731–736 (1984).
24. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. *Nucleic Acids Res.* **11**, 1475–1489 (1983).
25. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
26. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
27. Heidecker, G., Messing, J. & Gronenborn, B. *Gene* **10**, 69–73 (1980).

The correct activation of *Antennapedia* and bithorax complex genes requires the *fushi tarazu* gene

P. W. Ingham* & A. Martinez-Arias

Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge OX1 3PS, UK

In the *Drosophila* embryo the establishment and specification of metameric units depends upon the selective activation of the segmentation[1,2] and the homoeotic selector genes[3,4]. The former are necessary for establishing the appropriate number of metameric[2] or parasegmental[5] units, whereas the latter control the pathways of differentiation followed by particular parasegments[3,4]. Classical embryological manipulations[6,7] have shown that these processes must be closely coordinated during normal development. However, previous studies of pair-rule genes have led to the suggestion that the specification of segmental identity proceeds independently of the establishment of metameres as physical units[1,8]. These apparently conflicting perspectives can be reconciled by envisaging a common maternally derived positional information system which is independently interpreted by the components of both processes[6,9]. In the case of the partitioning process, the gap and pair-rule genes would be instrumental in translating this information, whereas the activation of the homeotic genes would be mediated via other intermediaries[4,10] (see ref. 9 for review). It is difficult to see, however, how such a system could ensure the precise regulation of the two types of genes implicit in the final differentiated pattern. This difficulty has led to the suggestion that the segmentation mechanism must define the precise boundaries of selector gene expression[11]. Here we confirm this suggestion and propose that the gene *fushi tarazu* plays a key role in this process, integrating the processes of metameric partitioning and regional specification in the *Drosophila* embryo.

The gene *fushi tarazu* (*ftz*)[2,12], a member of the pair-rule class, differs from the other members of the class in two interesting ways: it is located in the Antennapedia-complex (ANT-C)[13], one of two major clusters of selector genes, and it contains a class I homoeobox[14,15], which is present in and highly conserved between the ANT-C and bithorax-complex (BX-C) selector genes[14-16]. These characteristics have led to the suggestion that *ftz* is itself a homoeotic gene[17,18] which, in some unspecified way, regulates the expression of the ANT-C and BX-C genes[15]. Here we test these proposals and argue that *ftz* plays a dual role in the *Drosophila* blastoderm; in common with other pair-rule genes, it is necessary for the correct spatial regulation of segment-defining genes[19] for example *engrailed*, (*en*). In addition, we present evidence that it is necessary for the correct activation of homoeotic genes.

During normal embryogenesis the establishment of parasegmental units is first seen by the appearance of fourteen stripes of *en* activity along the antero-posterior axis of the early gastrula[19-21]. Shortly afterwards these stripes come to lie at the anterior margin of fourteen metameric units which can thus be defined as parasegments (PS)[5,22] (Fig. 1*a*). The specification of parasegments is illustrated at this stage by the distribution of selector gene products[5,16,23-26,28,29]. The *Ubx* proteins, for example, show a characteristic heterogeneous pattern between PS5 and PS13[27], defining a metameric repeat or 'Ubx-metamere'[28]; in the wild type these metameres coincide with parasegments and PS6 displays a characteristic higher intensity and more even distribution of proteins than other parasegments (see Fig. 1*b*).

In the extended germ band, *ftz*$^-$ embryos are partitioned into seven double-width parasegments[17] and lack all the even-numbered engrailed stripes[19] (Fig. 1*c*). At the same stage, embryos mutant for another pair-rule gene *odd paired* (*opa*), are similarly malpartitioned, though vestiges of the even numbered stripes of engrailed do remain (S. DiNardo and P. O'Farrell, personal communication; Fig. 1*e*). If the specification of parasegments is independent of their establishment, as previously suggested[1,8,9], the malpartitioning observed in *ftz*$^-$ and *opa*$^-$ embryos would be expected to generate parasegments of compound identity (see Fig. 2). To investigate this possibility we have analysed the distribution of *Ubx* proteins in *opa*$^-$ and *ftz*$^-$ embryos.

In embryos mutant for *opa*, the pattern of *Ubx* protein distribution in the double-sized units reveals them to be of composite character, the anterior half behaving as an odd-numbered parasegment and the posterior as an even-numbered parasegment. This is particularly clear in the case of the PS5/6 unit, in which the posterior half exhibits high levels of *Ubx* protein typical of the wild type PS6 (Fig. 1*f*). Each of the three more posterior double units displays defects in the *Ubx* pattern that are attributable to *opa* effects on segment polarity genes (unpublished data).

By contrast, the *Ubx* protein pattern in *ftz*$^-$ germ bands is quite different both from *opa* and wild type. The PS5/6 double unit has a staining pattern typical of the wild type PS5, except for a narrow strip of cells at the posterior border that show levels characteristic of a more posterior parasegment (but not PS6). The next double unit exhibits *Ubx* staining characteristic not of PS6 but of PS7-12, as so the next two double units, whereas the last double unit resembles PS13. The main feature of this pattern is that there are half the number of parasegments and half the number of Ubx-metameres. Both units coincide, and the *ftz*$^-$ embryos can be considered to consist of a chain of large odd-numbered parasegments: 3*, 5*, 7*, 9*, 11*, 13* (Fig. 1*d*).

Thus the two mutations *opa* and *ftz*, which result in similar malpartitioning of the germ band, show radically different patterns of *Ubx* expression within three hours of the blastoderm stage. The absence of PS6-specific *Ubx* staining in *ftz* germ bands cannot be accounted for by the elimination of PS6 cells, as little if any cell death has occurred in the ectoderm by this stage (unpublished observations). Rather, it suggests a failure in the specification of the metameric unit at blastoderm. We therefore monitored the initiation of *Ubx* expression in *ftz*$^-$ blastoderms.

An interesting and unexpected feature of the initial expression of *Ubx* in normal embryos is a transient pair-rule distribution of *Ubx* transcripts reminiscent of and in the same frame as that of *ftz*[11,23] (see Fig. 3). This discontinuous pattern is superimposed on a broader continuous distribution of transcripts. *Ubx* transcripts do accumulate in the *ftz*$^-$ blastoderm, but show a continuous distribution, with some modulation in the dorsal region (Fig. 3), but without the peak of accumulation ventrally or laterally in the anlage of PS6 that is the first sign of the pair-rule modulation in the wild type. Accumulation of transcript begins slightly posterior to the position of PS6 and probably reflects the wild-type early continuous distribution. At onset of gastrulation in wild-type embryos three further peaks of *Ubx* transcript[11,23] appear which correspond to PS8, PS10 and PS12 (Fig. 3); in *ftz*$^-$ embryos *Ubx* expression remains homogeneous at this stage. By the end of gastrulation, when the pair-rule modulation is normally still clearly visible, the pattern of *Ubx* expression reveals three unusually large and similar units reflecting the malpartitioning of the embryo and probably corresponding to the 7*, 9* and 11* parasegments described above (Fig. 3*f*).

To determine whether there is a requirement for *ftz*$^+$ in the activation of other selector genes we have studied the initiation of expression of the *Sex Combs Reduced* (*Scr*) and *Antennapedia* (*Antp*) loci in *ftz*$^-$ embryos. Expression of both genes also shows a transient pair-rule modulation in the late blastoderm[25,26]. In the wild type there is a strong peak of *Scr* expression in PS2

* Present address: Imperial Cancer Research Fund, Developmental Biology Unit, Department of Zoology, South Parks Road, Oxford, OX1 3PS, UK.

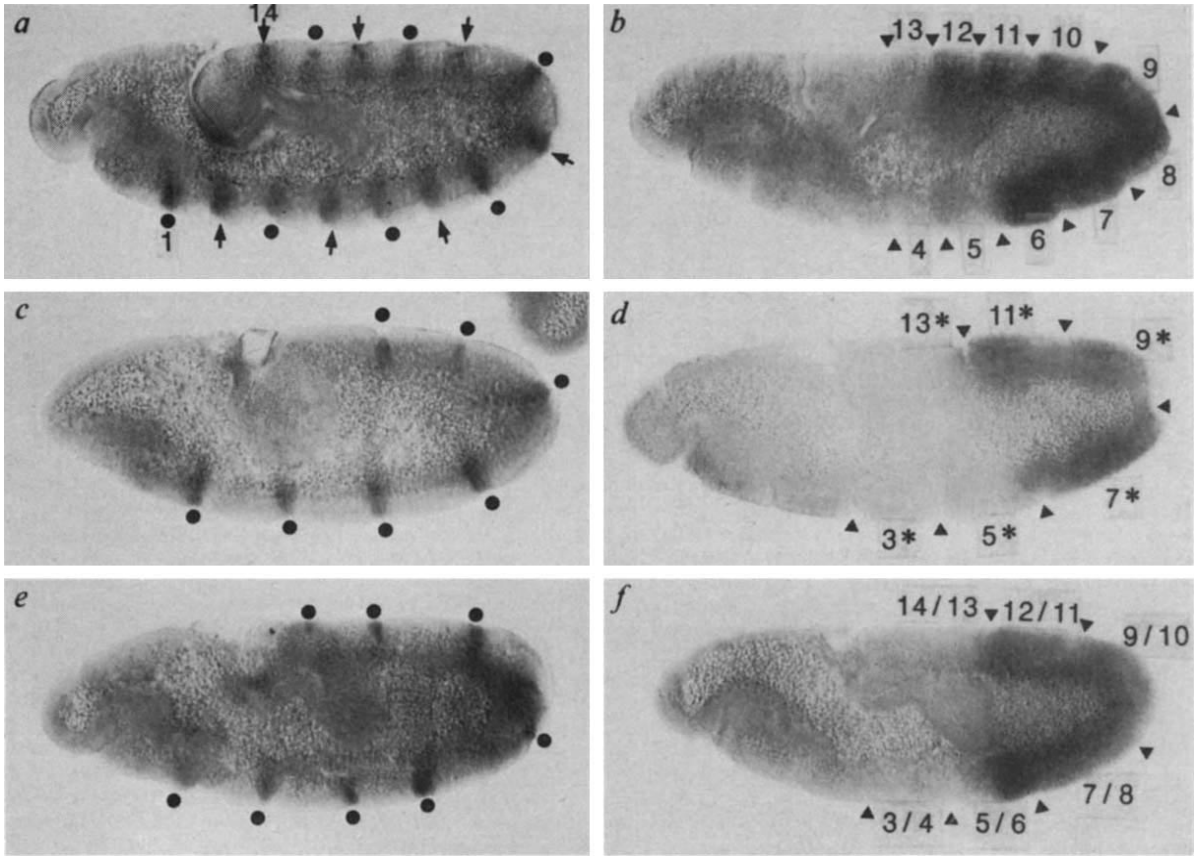


Fig. 1 Coupled and uncoupled partitioning and specification of parasegments visualized by *en* and *Ubx* protein distribution in wild-type and mutant embryos. *a*, Wild-type extended germ band showing bands of *en* protein (1-14), each of which is located at the anterior margin of the odd numbered (small dots) and even numbered (arrows) parasegmental units; *b*, wild type extended germ band showing typical *Ubx* protein distribution. The highest levels of staining are in PS6; PS7-12 show very similar patterns with higher levels posteriorly than anteriorly. PS5 has only limited expression around the tracheal pit and expression in PS13 is also restricted[23,27]; *c*, *ftz*$^-$ extended germ band stained with *en* antibody. The embryo is subdivided into seven double sized units and only the odd-numbered *en* stripes are present; these are of approximately the same width as wild type, in agreement with previous transcriptional data[19]. *d*, embryo similar to that shown in *c* stained with antibody for *Ubx*. There is no staining characteristic of the wild type PS6. The most anterior unit which expresses *Ubx* appears as an enlarged PS5, with the exception of a few cells at its posterior margin which show a higher level of staining. The following three units have patterns typical of single parasegments, which would correspond to 7*, 9* and 11*; *e*, *opa*$^-$ extended germ band stained with *en* antibody. Like the *ftz*$^-$ germ band, it is composed of half the normal number of double width parasegmental units, only of the odd-numbered *en* bands are present. However, vestiges of even-numbered bands are often visible, especially in the mesoderm. *f*, Embryo similar to that shown in *e*. A single segment periodic pattern of *Ubx* staining similar to that of wild type is superimposed on the malpartitioned embryo which is composed of half the wild-type number of double segment units. Note especially that the PS5-6 unit shows staining characteristic of the wild type PS5 and PS6. The deviations from wild type in the more posterior parasegments can be explained by the effects of segment polarity genes on *Ubx* (unpublished observations).

Methods. Embryos were stained with monoclonal antibodies against *en*[20] and *Ubx*[41]. The staining procedure is similar to that described elsewhere[27], except that the second antibody was biotinylated, allowing the use of peroxidase conjugated to a biotin-avidin system (Vectastain kit; Vector laboratories). The stain was visualized with diamino benzidine according to instructions from the suppliers. In these and subsequent experiments, we used three alleles of *ftz*: a deficiency *Df(3R)4Scb*, a point mutant, *ftz*^{9H34} (ref. 42), and an insertion *ftz*^{W20} (ref. 18). In the case of *odd paired* a single allele *opa*^{5H97} (ref. 42) was used.

and six weaker peaks corresponding to even-numbered parasegments. This initial pattern of *Scr* expression is completely abolished in *ftz*$^-$ embryos (Fig. 4*c*).

Expression of *Antp* from the P2 promoter is first detectable in a four-cell-wide peak which corresponds in position to the primordium of PS4 in the late blastoderm[26], this peak is not present in *ftz*$^-$ embryos at the same stage (Fig. 4*b*). Transcripts from the P2 promoter only appear in *ftz*$^-$ embryos during late gastrulation, when they are restricted to the presumptive ectoderm in a region corresponding to PS3 and PS4 (data not shown). This is similar to *Antp* expression in wild-type PS3[26] and is likely to correspond to PS3*. The pattern of *Antp* expression from the P1 promoter is not altered in *ftz*$^-$ embryos (unpublished observations).

These results suggest that *ftz* plays a fundamental role in differentiating between the even- and odd-numbered parasegments by modulating the activation of the selector genes in the blastoderm. Although the unique identity of each parasegment is controlled by selector genes, the genes themselves are generally expressed in more than one parasegment[16,23-29]. A striking feature of the presumptive gnathal and thoracic region of the embryo at blastoderm (and the central nervous system during later embryonic stages), is that the even-numbered parasegments are the primary sites of expression of different selector genes[23-26]. Thus *Scr* is expressed at highest levels in PS2, *Antp* in PS4 and *Ubx* in PS6; mutations in each of these genes transform their preferred parasegments but not the other even-numbered ones. By contrast, both genetic analysis and expression patterns suggest that the odd-numbered parasegments depend on more than one selector gene for their correct development. Different parasegments thus come to be distinguished both by different combinations and different levels of selector gene activity. Both PS4 and PS5, for instance, express *Antp*, but this expression differs between the two, both in spatial pattern and amount. This

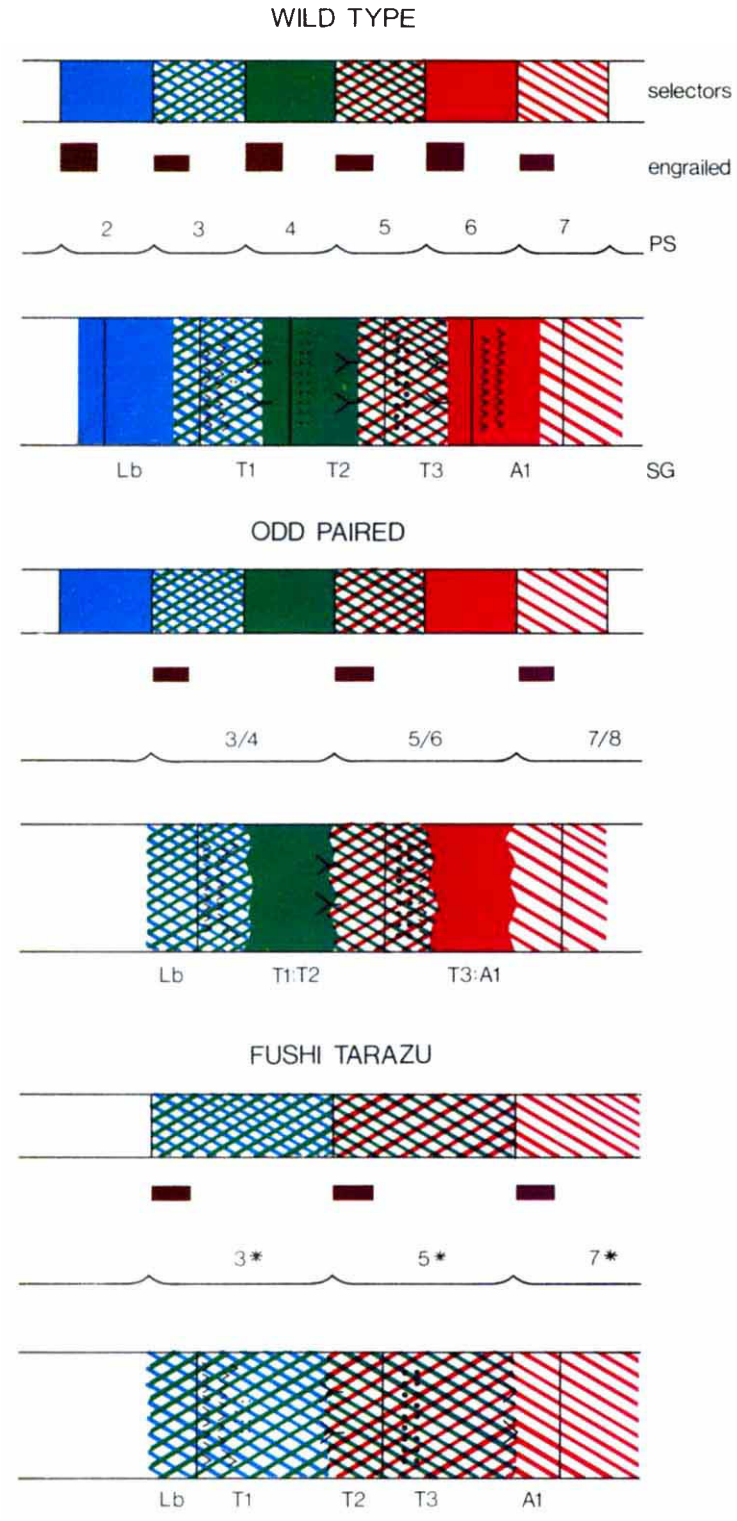


Fig. 2 Schematic representations of the embryonic development in wild type and each of the mutants. Colours indicate activity of selector genes: blue, *Scr*; green, *Antp* and red, *Ubx*. Cells in even-numbered parasegments express a single selector gene in a rather homogeneous pattern. Odd numbered parasegments display complicated patterns with more than one selector gene expressed at high levels in areas not demarcated by compartment boundaries[23-26] (indicated by hatching). Notice how segments are derived from parasegments by shifting the position of the grooves; this occurs during muscle attachment after germ band shortening[5]. In the wild type, segments are thus composed of two adjacent parasegments. In the mutants, this is also true but the selector gene distribution in them results in enlarged segments with rather different identities from wild type. The two different mutant situations illustrated will result in different cuticular phenotypes. In the case of *opa*, where partition and specification are uncoupled, anterior compartments are composed of cells of two different identities; this in turn generates larval cuticular denticle bands of composite segmental character, as observed in *opa*$^-$ larvae[2]. The wavy boundaries represent the observed variability of the phenotype. In the case of *ftz*, where partition and specification are coupled, anterior compartments have a single identity; the final phenotype thus corresponds to the precise elimination of alternate parasegments[2]. Notice that our observations and interpretations contrast sharply with former interpretations of these phenotypes[2], which suggest that the phenotypic differences between *opa* and *ftz* directly reflect a requirement in different but overlapping regions; this requirement would correlate with the site of expression of the gene. In our view, the phenotypes reflect the fundamental differences in the functions rather than the spatial depolyment of both genes.

difference is maintained by *Ubx*, as in the absence of *Ubx*, PS5 (and PS6) now develop like PS4 and express *Antp* accordingly[30,31]. These differences between *Antp* expression in the primordia of PS4 and PS5 are, however, apparent in the late blastoderm stage, long before *Ubx* product could possibly act and indeed in the absence of the *Ubx* gene[29]. It follows that the parasegmental modulation of *Antp* must be established independently of *Ubx* and only subsequently maintained by interactions between selector genes. Similarly *Scr*, which is required and expressed in PS2 and PS3, is expressed at high levels only in the PS2 primordium at blastoderm, whereas *Ubx* is expressed at high levels in the PS6 primordium at the same time.

The peaks of selector gene activity thus seem to differentiate between the even-numbered and odd-numbered parasegmental domains from the onset of their expression. The results presented here show that the patterns of *Scr*, *Antp* and *Ubx* all depend on *ftz* activity; in the absence of *ftz* products none of the peaks are established. This does not prevent the subsequent expression of any of these genes, but the absence of the initial modulation is reflected in the later failure to establish differences between PS3 and 4, 5 and 6 and so on in *ftz*$^-$ embryos. A prediction of this interpretation is that expression of *Scr* and *Antp* extends further posteriorly than normal, so that *Scr* now becomes expressed in cells which would normally have given rise to PS4,

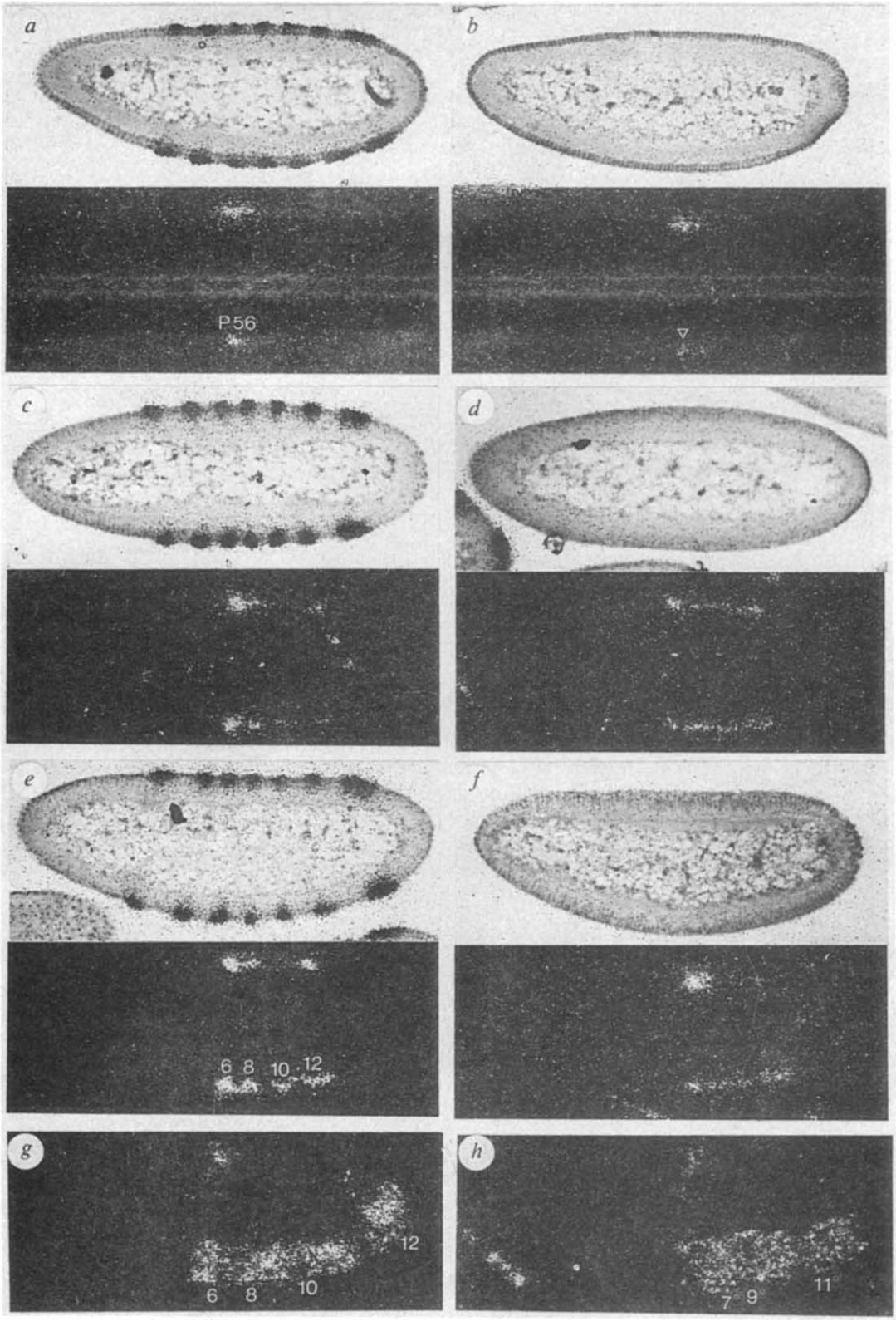


Fig. 3 Initiation of *Ubx* transcription in wild type and *ftz*⁻ embryos. Panels *a–f* show bright field images of sections hybridized with ^{35}S-labelled *ftz* probe revealing the *ftz* embryos and dark field images of adjacent sections hybridized with ^{35}S-labelled *Ubx* probe. *a*, Wild type; a peak of *Ubx* transcript is first detectable dorsally and ventrally in a region corresponding to PS6; *b*, *ftz*⁻ embryos at the same stage; there is no clear peak of transcript ventrally and laterally (arrowed). *c*, Slightly older wild type embryos show rather high levels of *Ubx* transcript in PS6, with a low background of transcription extending posteriorly; *d*, no such peak is observed laterally or ventrally in *ftz*⁻ embryos. *Ubx* transcript is however detected at low levels in a broad domain from 20% to 50% egg length. *e*, With the onset of gastrulation in wild type three further peaks appear[11,23], corresponding to PS8, 10 and 12; *f*, in *ftz*⁻ embryos, however, the expression ventrally remains more or less uniform. *g*, *h*, dark field images of *Ubx* labelling in late gastrulae. In wild type the pair-rule modulation persists, the label revealing the presumptive parasegments 6, 8, 10 and 12. In *ftz*⁻ embryos, by contrast, the label reveals three adjacent units which probably correspond to the 7*, 9* and 11* parasegments shown in Fig. 3. The *in situ* methods used are similar to those described elsewhere[43]. The *ftz* probe spans all the *ftz* coding region[43]; the *Ubx* probe corresponds to the A122 fragment[23], and in Northern blots detects a 4.7-kilobase (kb) RNA.

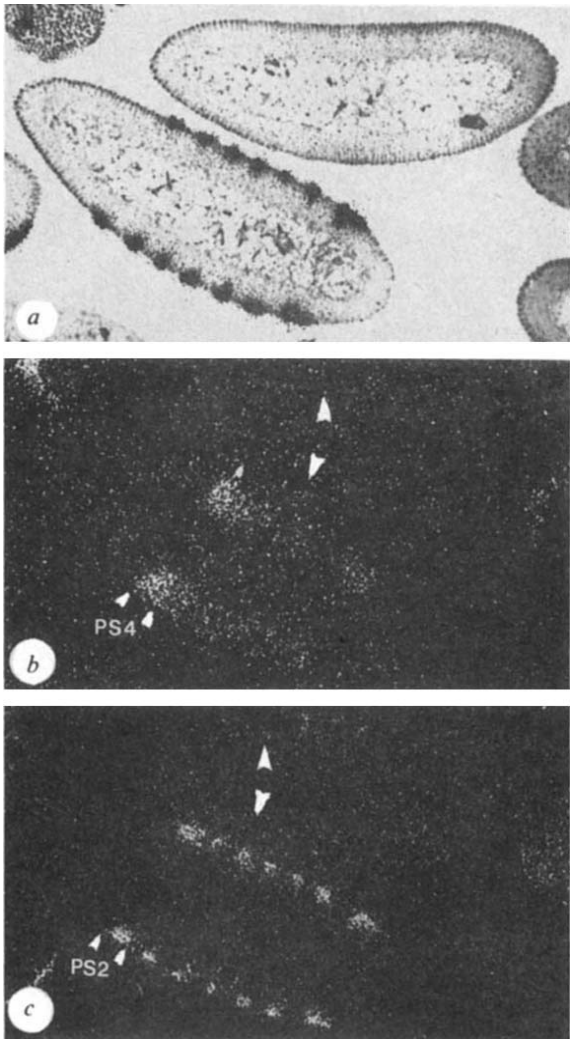


Fig. 4 Expression of *ftz*, *Antp* and *Scr* in wild type and ftz^{W20} blastoderms. *a*, Saggital sections through wild type (lower) and ftz^{W20} (upper) late stage 14 embryos hybridized with ^{35}S-labelled *ftz* probe. The wild type shows the typical pattern of seven discrete bands of transcript accumulation. The mutant shows very low levels of transcript accumulation which show a weak pair-rule modulation. *b*, Dark field image of adjacent section to *a* hybridized with a ^{35}S-labelled probe specific for the *Antp* P2 transcript. This reveals accumulation of high levels of transcript in the cells of the PS4 primordium in the wild type, but not in the same region of the ftz^{W20} embryo (arrowed). *c*, Dark-field image of adjacent section to *a* hybridized with a ^{35}S-labelled probe derived from the 5′ end of the *Scr* transcription unit. In the wild type, *Scr* transcript accumulates in seven stripes which correspond in their position and size to those of *ftz* transcript. Transcript accumulation is higher in the most anterior of these, which corresponds to the primordium of PS2. In the ftz^{W20} embryo there is no detectable expression of *Scr* at this stage, either in PS2 (arrowed) or more posteriorly. The *Antp* probe is a fragment from the proximal part of the gene; at blastoderm it specifically detects transcripts from the P2 promoter[26]. Methods as in Fig. 3.

whilst *Antp* is expressed in cells which would normally make PS6. This implies that in the wild type the peaks of expression serve to differentiate between parasegments by establishing specific levels or states of selector gene activity in the cells of the even-numbered parasegments. These states subsequently make these cells refractory to the expression of the more anteriorly expressed selector genes. This close linkage of the selector gene pattern with *ftz* expression is emphasized by the finding that it is maintained in embryos mutant for the gap genes

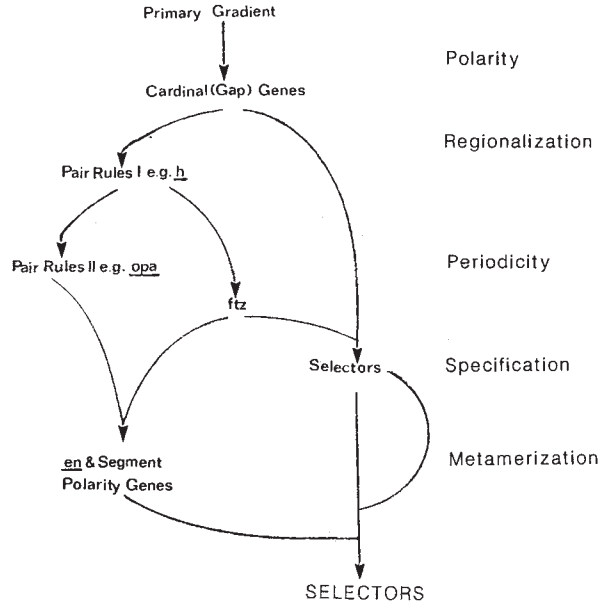


Fig. 5 A summary of the hierarchial interactions which lead to the integrated partitioning and specification of the *Drosophila* embryo. Arrows indicate the regulatory relationships between different classes of genes. The processes proposed to be accomplished by each step are indicated on the right. A primary gradient of maternally derived morphogen(s) establishes 'polarity' and, as the blastoderm begins to form, instructs the deployment of the cardinal or gap genes[40]. This results in the 'regionalization' of the embryo, that is, the establishment of the broad domains of potential selector gene expression. The cardinal genes also act to establish 'periodicity' by organizing the expression of a subset of the pair-rule genes, exemplified by *hairy* (*h*)[32]. The function of this class is to pattern *ftz*, and also, we surmize, other pair-rule genes such as *opa*. This step-by-step subdivision of the embryo is translated into an array of parasegments that are uniquely specified, primarily by the activation of both *en* and the ANT-C and BX-C genes by *ftz*. The 'specification' and 'metamerization' of parasegments takes place as the blastoderm forms into cells. Later, the interactions indicated elaborate the pattern of selectors to that of the first instar larva. This elaborated pattern of selector gene expression is represented by capital letters.

Kr, *kni* and *hb* all of which alter the pattern of *ftz* expression[32,33].

Previous analyses have demonstrated a functional non-equivalence of the pair-rule genes, some being required for normal expression of *ftz* whilst others are not[19,34]. These findings led to the proposition that parasegments are established by a hierarchy of interactions between pair-rule genes resulting in the establishment of the pattern of *en* expression[19]. We propose a further subdivision of the pair-rule class distinguishing between those genes that are responsible only for the partitioning process (for example, *opa*) and *ftz*, which acts both to partition and to specify the cells of the blastoderm. Thus, the processes of partitioning and specification of the *Drosophila* embryo are integrated through a hierarchical network of interactions between the different segmentation genes (see Fig. 5).

As *ftz* itself is expressed in a reiterated pattern throughout the presumptive germ band, it is clear that it cannot alone account for the differential activation of the selector genes. There must therefore be specific activators which allow *ftz* to discriminate between the selector genes in different regions of the blastoderm; candidates for such activators are the zygotic gap or cardinal genes[1,2] whose products are distributed in broad though restricted domains[35] and whose absence has been shown to affect selector gene expression[27,32,33]. Consequently, understanding the correlation between the metameric pattern and selector gene expression demands a consideration of the possible molecular interactions between *ftz* and the cardinal gene products.

The data available[32,33] suggest that an interaction between *ftz* and cardinal gene products is responsible for the activation of selector genes, the *ftz* product providing the precise spatial specificity and the gap gene products the promoter specificity. One possibility is that *ftz* and the cardinal products are activators; the specificity of selector gene expression would then result from a simple combination of their activities. An alternative envisions a maternally encoded repressor common to all selector genes and bound at the sites of cardinal gene activity. The role of *ftz* would be to compete out this repressor and thus make the underlying pattern of cardinal gene activity available to selector genes; in this possibility the role of *ftz* in activation, though crucial, would be indirect.

Our model contrasts with those previously proposed in which a primary gradient, maternally encoded, is directly responsible for the activation of the homoeotic selector genes[4,39]. Although there is good evidence for the existence of such gradients[6,36-38], it seems unlikely that they can have the instructive capacity required for the short range detailed patterning that precisely orders selector genes in the zygote. They can, however, as suggested by Meinhardt[40], have instructive capacity in broader, less precisely defined ranges, and may activate the cardinal genes. Thus, the ultimate effect of such gradients on the selector genes would operate through the aforementioned hierarchy (Fig. 5) and ultimately rely on their indirect effect on *ftz*.

We thank M. Akam for encouragement and many critical discussions, C. Cabrera, D. Hogness, K. Howard, P. Lawrence and K. Nasmyth for interest and discussions, C. Nüsslein-Volhard, G. Jürgens and T. Kaufman for mutant stocks and M. Scott for ANT-C probes. We thank M. Eggo, B. Cross and A. Donoughue for typing. A.M.-A. was supported by an EMBO long-term fellowship.

Note added in proof: Results compatible with our conclusions have recently been reported by Duncan, I. M. *Cell*, **47**, 297-309 (1986).

Received 6 October; accepted 17 November 1986.

1. Nüsslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795-801 (1980).
2. Nüsslein-Volhard, C., Wieschaus, E. & Jürgens, G. *Verh. dt. zool. Ges.* 91-104 (1982).
3. Garcia-Bellido, A. *Ciba Fdn Symp.* **29**, 161-182 (1975).
4. Lewis, E. B. *Nature* **276**, 565-570 (1978).
5. Martinez-Arias, A. & Lawrence, P. A. *Nature* **313**, 639-642 (1985).
6. Sander, K. *Adv. Insect Physiol.* **12**, 125-238 (1976).
7. Newman, J. M. & Schubiger, G. *Devl Biol.* **79**, 128-138.
8. Sander, K., Lohs-Schardin, M. & Baumann, N. *Nature* **287**, 841-843 (1980).
9. Lawrence, P. A. *Cell* **26**, 3-10 (1981).
10. Struhl, G. *Nature* **293**, 36-41 (1981).
11. Akam, M. E. *Phil. Trans. R. Soc.* **B312**, 179-187 (1985).
12. Wakimoto, B. T. & Kaufman, T. C. *Devl Biol.* **81**, 51-64 (1981).
13. Lewis, R. A., Kaufman, T. C., Denell, R. E. & Tallerico, P. *Genetics* **95**, 367-381 (1980).
14. Scott, M. P. & Weiner, A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4115-4119 (1984).
15. Laughon, A. & Scott, M. P. *Nature* **310**, 25-31 (1984).
16. McGinnis, W., Levine, M., Hafen, E., Kuroiwa, A. & Gehring, W. J. *Nature* **308**, 428-433 (1984).
17. Wakimoto, B. T., Turner, I. R. & Kaufman, T. C. *Devl Biol.* **102**, 147-172 (1984).
18. Weiner, A. J., Scott, M. P. & Kaufman, T. C. *Cell* **37**, 843-851 (1984).
19. Howard, K. R. & Ingham, P. W. *Cell* **44**, 949-957 (1986).
20. DiNardo, S., Kuner, J. M., Theis, J. & O'Farrell, P. H. *Cell* **43**, 59-69 (1985).
21. Kornberg, T., Siden, I., O'Farrell, P. & Simon, M. *Cell* **40**, 45-53 (1985).
22. Ingham, P. W., Martinez-Arias, A., Lawrence, P. A. & Howard, K. R. *Nature* **317**, 634-636 (1985).
23. Akam, M. E. & Martinez-Arias, A. *EMBO J.* **4**, 1689-1700 (1985).
24. Martinez-Arias, A. *EMBO J.* **5**, 135-141 (1985).
25. Martinez-Arias, A., Ingham, P. W., Scott, M. P. & Akam, M. E. *Development* (submitted).
26. Martinez-Arias, A., Scott, M. P. & Akam, M. E. (in preparation).
27. White, R. A. H. & Lehmann, R. *Cell* **47**, 311-321 (1986).
28. White, R. A. H. & Wilcox, M. *EMBO J.* **4**, 2035-2044 (1985).
29. Harding, K., Weeden, C., McGinnis, W. & Levine, M. *Science* **229**, 1236-1242 (1985).
30. Hafen, E., Levine, M. & Gehring, W. J. *Nature* **307**, 287-289 (1984).
31. Carroll, S., Laymon, R. A., McCutcheon, A., Riley, P. D. & Scott, M. P. *Cell* **47**, 113-122 (1986).
32. Ingham, P. W., Ish-Horowicz, D. & Howard, K. R. *EMBO J.* **5**, 1659-1665 (1986).
33. Martinez-Arias, A., Akam, M. E. & Ingham, P. W. (in preparation).
34. Carroll, S. B. & Scott, M. P. *Cell* **45**, 113-126 (1986).
35. Knipple, D., Seifert, E., Rosenberg, U., Preiss, A. & Jäckle, H. *Nature* **317**, 40-44 (1985).
36. Lehman, R. thesis, Tübingen Univ. (1985).
37. Mlodzik, M., Fjose, A. & Gehring, W. J. *EMBO J.* **4**, 2961-2969 (1985).
38. Schüpbach, T. & Wieschaus, E. *Wilhelm Roux Arch. dev. Biol.* **195**, 302-317 (1986).
39. Capdevila, M. P. & Garcia-Bellido, A. *Wilhelm Roux Arch. dev. Biol.* **190**, 1339-350 (1981).
40. Meinhardt, H. *J. Cell Sci. Suppl.* **4**, 357-387 (1986).
41. White, R. A. H. & Wilcox, M. *Cell* **39**, 163-171 (1984).
42. Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C. & Kluding, H. *Wilhelm Roux Arch. dev. Biol.* **193**, 283-296 (1984).
43. Ingham, P. W., Howard, K. R. & Ish-Horowicz, D. *Nature* **318**, 439-445 (1985).

Does nuclear integration of mitochondrial sequences occur during senescence in *Podospora*?

France Koll

Centre de Génétique Moléculaire, Centre National de la Recherche Scientifique, 91190 Gif-sur-Yvette, France

Observations in various organisms[1-3] suggest that the transfer of mitochondrial DNA sequences to the nucleus has occurred in the course of evolution. Wright and Cummings have reported[4], on the basis of hybridization experiments, that there is such a transposition in the fungus *Podospora anserina* during senescence[5] (arrest of vegetative growth) which is accompanied by amplification of specific mitochondrial sequences[6-9]. They suggested that this transposition could explain senescence, and, as it could be regularly observed under laboratory conditions, senescence in *Podospora* could constitute a model system to study the molecular mechanisms of the transfer of DNA sequences from mitochondria to the nucleus. We report here experiments similar to those of Wright and Cummings[4] but include additional critical controls. Our results do not confirm the data of Wright and Cummings; the only positive signals we observed after hybridization of mitochondrial DNA sequences on nuclear DNAs can be attributed to contaminating DNA from mitochondria or plasmids.

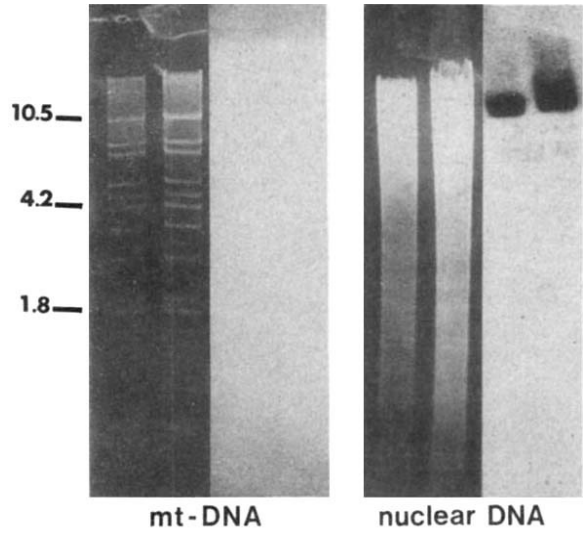


Fig. 1 Detection of a unique nuclear sequence. *a*, Wild-type strain; *b*, *mex1* strain. Nuclear and mt DNAs obtained from a total DNA extract[15] after separation on diamidino-phenyl-indole (DAPI)-CsCl buoyant density gradient[7] (3 or 4 cycles). The 1.4-kb nuclear sequence used as probe contained the suppressor gene *su4* of *Podospora* previously selected from a genomic library[19] and was nick-translated[20] with [α-^{32}P]dATP and [α-^{32}P]dCTP (specific activity $\sim 5\times10^7$ c.p.m. μg^{-1}), hybridized as already described[15] with nitrocellulose filters on which restriction digests[21] of the DNAs were transferred according to Southern[22]. Shown here are the autoradiograms observed after hybridization of the 1.4-kb nuclear sequence on the *Bam*HI digestion of ~ 5 μg nuclear and ~ 0.5 μg mt-DNA from a young wild-type strain and from the long-lived mutant *mex1* (nuclear DNA complexity: 2×10^7 base pairs (bp); mt-DNA complexity: 9×10^4 bp). The electrophoretic patterns of the DNAs from the wild-type strain are represented in *a* and from the *mex1* strain in *b*. The autoradiograms (2 days exposure at room temperature) are represented in *a*' and *b*'. We detect an 11-kb sequence similar to the probe in the nuclear DNA of each strain. The probe does not hybridize on the mt-DNA. Sizes in kb.

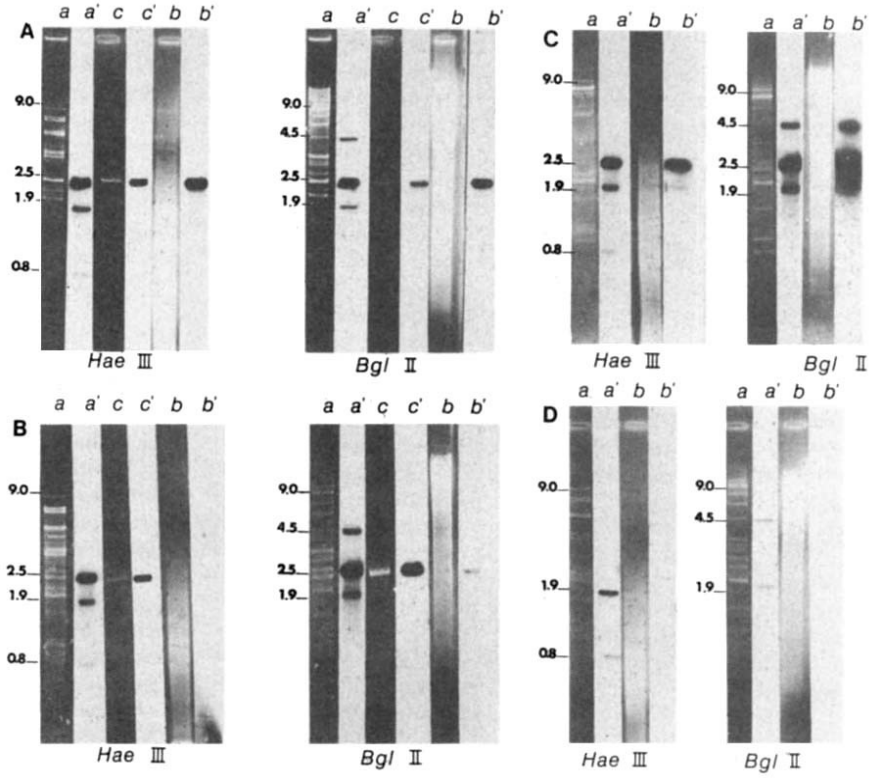


Fig. 2 Hybridization of the α mitochondrial intron on the different DNAs extracted from three independent senescent cultures (A, B, C) and one young *s mt+* (D). Nuclear DNA (density, 1.712 g cm^{-3}), α-free sen-DNA molecules (density, 1.700 g cm^{-3}) and mt DNA (density, 1.694 g cm^{-3}) of each senescent culture were purified by at least three cycles of DAPI-CsCl ultracentrifugation of total DNA extract. Each DNA was digested by *Hae*III and *Bgl*II, electrophoresed on agarose gels, transferred to nitrocellulose filters and hybridized with a ^{32}P nick-translated α purified sequence. The α sequence was purified from the plasmid in which it was integrated by *Bam*HI digestion, electrophoresis on acrylamide gel and electroelution[15]. For both enzymes (*Hae*III, *Bgl*II), the electrophoretic pattern of mt-DNA, α-free sen-DNA molecules and nuclear DNA from a given culture are represented respectively in lanes *a*, *b*, *c* after staining by ethidium bromide (α-sen-DNA extracted from the culture C is not represented). Corresponding autoradiograms after hybridization are shown in lanes *a'*, *b'* and *c'*. For mt-DNA and α-sen-DNA, autoradiograms were exposed for 24 h at room temperature; for nuclear DNAs, they were observed after 4–5 days at −70 °C using two intensifying screens. D, Control experiment showing hybridization of the α sequence on *Hae*III and *Bgl*II digests and corresponding autoradiograms of mt-DNA (*a* and *a'*) and nuclear (*b* and *b'*) DNAs from a young *s mt+* wild type strain. The α mt-sequence used as probe is 2.5-kb long and contains only one site for both enzymes *Hae*III and *Bgl*II. Two fragments (1.9 and 0.8 kb after *Hae*III digest; 4.5 and 1.9 kb after *Bgl*II digest), corresponding to the α sequence integrated in the mitochondrial genome, are labelled after hybridization of the probe on the young mt-DNA (D, lanes *a'* and *b'*). The free α-sen-DNA molecules present in the senescent cultures are circular molecules constituted of tandemly repeated α sequence. Consequently, only one 2.5-kb fragment corresponding to linearized α sequence is observed after digestion of this α-*sen*-DNA by *Hae*III and *Bgl*II (A and B, lanes *c*) and hybridized with the probe (A and B, lanes *c'*). The three labelled bands observed on *Hae*III and *Bgl*II digestion of senescent mt-DNA (parts A, B and C, lanes *a'*) correspond to hybridization of the probe on the α sequence integrated in the mitochondrial genome (1.9- and 0.8-kb fragments on *Hae*III digests; 4.5- and 1.9-kb fragments on *Bgl*II digests) and on linearized free α-sen-DNA molecules always associated with it.

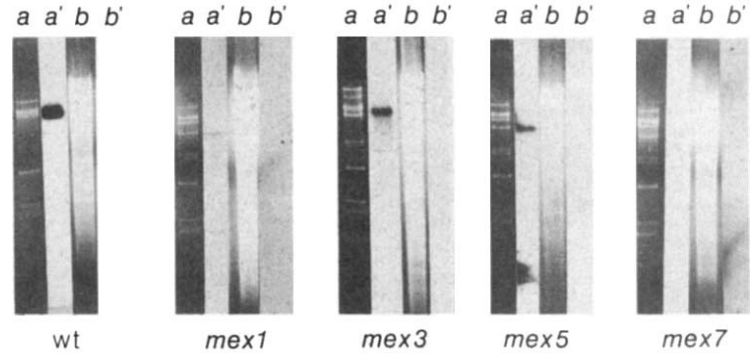


Fig. 3 Hybridization of the α sequence on the nuclear DNA of the mitochondrial long-lived mutants *mex1*, *mex3*, *mex5*, *mex7* and of *s mt+* wild-type strain (wt). The structure of the mt DNA of each mutant has already been described[14,15]: this DNA is deleted for most of the α mitochondrial intron in *mex1*, *mex5* and *mex7* mutants and it is not modified in this sequence in *mex3*. Sequencing data indicate that the *mex5* mt-DNA has retained 387 bp of the α sequence whereas only 32 bp are conserved in *mex1* and *mex7* mt-DNA. The nuclear DNA and mt-DNA of each strain were purified from total DNA extracts after DAPI-CsCl centrifugation. After digestion by *Eco*RI and gel electrophoresis, each nuclear and mt-DNA was transferred to nitrocellulose filters and hybridized with labelled α sequence (see Fig. 2 legend) as in Fig. 1. The electrophoretic patterns of the mt-DNA and of the nuclear DNA are represented in lanes *a* and *b*, respectively, and the corresponding autoradiograms in lanes *a'* and *b'*. These autoradiograms were observed after 4–5 days of exposure of the filters at −70 °C except for the *mex3* and wild-type mt-DNA (24 h at room temperature). Note that the specific activity of the probe and the hybridization conditions were such that the *Eco*RI fragment of *mex1* and *mex7* mt-DNA, which contains only 32 bp of the α sequence could be detected. In these conditions no hybridization was observed on the nuclear DNAs of these mutants.

In *Podospora*, senescence[5] is assumed to be caused by the amplification of some mitochondrial (mt) DNA sequences, called sen-DNAs[7–11]. The most frequently amplified sequence, α-sen-DNA, corresponds exactly to the first intron of the mt-gene encoding the first subunit of cytochrome oxidase[12,13]. We looked for integration of this α sequence into the nuclear genome of three senescent cultures of the wild-type strain *smt+* and in that of four long-lived mitochondrial mutants (*mex1*, *mex3*, *mex5* and *mex7*)[14,15]. Note that *mex1*, isolated in our laboratory[10], was reported by Wright and Cummings[4] to contain the α sequence integrated within its nuclear genome.

The nuclear DNA of these different strains was hybridized with the α sequence (Fig. 2) in conditions allowing us to detect a single copy of a nuclear sequence (see Fig. 1 legend). To exclude the possibility that the putative labelled bands could result from hybridization of the probe on contaminating DNAs (mt-DNA or free α-sen-DNA molecules which are found as bands between those of nuclear and mt-DNA in a buoyant density caesium chloride gradient[6,7]), hybridizations were simultaneously performed on mt-DNA and purified α-sen-DNA. Figure 2 shows the results of the hybridization of α sequence on the *Hae*III- and *Bgl*II-restricted nuclear DNA from three independent senescent cultures, A, B and C. Both enzymes have one site in the α sequence. Consequently, we expected to find two or three labelled bands if the α sequence is integrated in the nuclear genome as a single copy or tandem multicopy,

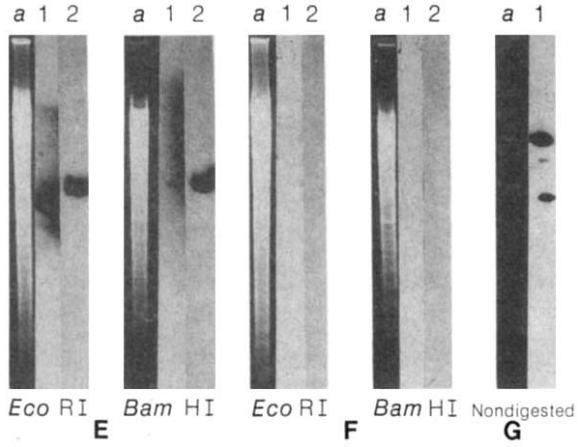


Fig. 4 Hybridization of the purified α sequence and pBR322 DNA on two nuclear DNAs obtained from two independent cultures of the *mex1* mutant. Each nuclear DNA was digested by *Eco*RI and *Bam*HI, transferred to nitrocellulose filter and hybridized with purified α sequence (see Fig. 2) and separately with pBR322 DNA (extracted according to Maniatis *et al.*[20]) used as radioactive probes. For each nuclear DNA samples (E and F), the electrophoretic patterns of the *Bam*HI and *Eco*RI digestions are represented in lanes *a*. The autoradiograms (4–5 days at −70 °C) observed after hybridization of the α sequence and pBR322 DNA are represented in lanes 1 and 2, respectively. On the nuclear DNA E, no hybridization is observed. On the nuclear DNA D, one labelled band is observed with both probes at the same position whatever the enzymes used (*Bam*HI or *Eco*RI). This suggests that these labelled bands result from hybridization of pBR322 or derivative DNA contaminating both the nuclear DNA E and the 'purified' α sequence. Hybridization between pBR322 DNA and purified α sequence, represented in G, confirmed that this purified DNA is contaminated by pBR322. Nondigested pBR322 DNA was electrophoresed, transferred to nitrocellulose and hybridized with the purified α sequence. The electrophoretic pattern of nondigested pBR322 DNA is presented in lane *a* and the corresponding autoradiogram in lane 1. The three forms of pBR322 DNA are detected by the probe.

respectively. In the latter case, one of these bands must be 2.5 kilobases (kb) long as this is the length of the α sequence.

For cultures A and B, we observed a single 2.5-kb labelled band, which we interpret as hybridization on free α-sen-DNA molecules contaminating the nuclear DNAs. For culture C, three labelled bands are observed, one at 2.5 kb, suggesting that head-to-tail copies of α are integrated, and the others, which would correspond to the sequences flanking α, migrate as the mt-fragments that hybridize with the α sequence. Our conclusion, confirmed by using other enzymes (data not shown), is that this nuclear DNA is contaminated both by mt-DNA and by free α-sen-DNA. Such contamination is difficult to avoid, even after three or four cycles of purification on CsCl gradients, because of the small differences in density of the three molecular species recovered from senescent cultures.

Figure 3 shows the results of hybridization of *Eco*RI-restricted nuclear and mt-DNAs of the mutants *mex1*, *mex3*, *mex5* and *mex7* with the α sequence which has no *Eco*RI site. Nuclear DNAs did not hybridize with the probe.

Figure 4 represents hybridization of α and pBR322 DNA probes on two different nuclear DNA extracts (E, F) of *mex1*. We found one labelled band at the same position with both probes on nuclear DNA (E) but no label on nuclear DNA (F). The nuclear DNA (E) was probably contaminated by pBR322 or derivative DNA. Contamination by plasmid DNA is fairly common; Shmookleir Reis *et al.*[16] have pointed out the danger of systematic microbial contamination of DNA samples. Indeed, the presence of plasmidic DNA, both in the DNA used as a radioactive probe (even after two cycles of electrophoretic purification[17]) and in the DNA samples tested, led to the appearance of labelled bands that could be misinterpreted in the absence of appropriate controls such as hybridization using the vector alone as probe.

Our experiments, repeated several times, failed to demonstrate any similarity between the α sequence and the nuclear DNA, from either senescent cultures or long-lived mutants. This contradicts the results of Wright and Cummings[4], who interpreted their hybridization experiments, for senescent cultures as well as for *mex1*, as positive. Two hypotheses can be proposed to explain these discrepancies: (1) the nuclear DNA from the subclones studied by Wright and Cummings contains mitochondrial sequences whereas the nuclear DNA from the subclones we have tested does not; and (2) whichever subclones studied, the nuclear DNA has no similarity with the α-mitochondrial sequence, the additional controls included in our experiments explaining their results as being caused by contamination by DNA from mitochondria or plasmids. In the absence of sequencing data demonstrating that a nuclear DNA actually flanks the α sequence, we favour the second hypothesis.

Our data show that a transposition of α sequence to the nucleus is not systematically associated with the senescence process and therefore cannot explain the arrest of growth in *Podospora*. This conclusion is also consistent with the genetic data[18] collected previously which shows that the senescent state is maternally inherited.

I thank L. Belcour and C. Vierny for help and advice, J. Beisson and L. Sperling for critical reading of the manuscript; A. Sainsard and O. Begel for help; and Y. Brygoo and R. Debuchy for the gift of the plasmid containing the *su4* nuclear gene of *Podospora*. I acknowledge the receipt of an ADRC fellowship. This work was supported by a grant from the Centre National de la Recherche Scientifique (ATP Microbiologie).

Received 3 September; accepted 28 October 1986.

1. Farelly, F. & Butow, R. A. *Nature* **301**, 296-301 (1983).
2. van den Boogaart, P., Samello, J. & Agsterribb, E. *Nature* **298**, 187-189 (1982).
3. Jacobs, H. T. *et al. J. molec. Biol.* **165**, 609-632 (1983).
4. Wright, R. M. & Cummings, D. J. *Nature* **302**, 86-88 (1983).
5. Rizet, G. *C.r. hebd. Séanc. Acad. Sci. Paris* **237**, 838-840 (1953).
6. Stahl, U., Lemke, P. A., Tudzynski, P., Kück, U. & Esser, K. *Molec. gen. Genet.* **162**, 341-343 (1978).
7. Cummings, D. J., Belcour, L. & Grandchamp, C. *Molec. gen. Genet.* **171**, 239-250 (1979).
8. Jamet-Vierny, C., Begel, O. & Belcour, L. *Cell* **21**, 189-194 (1980).
9. Cummings, D. J., Laping, J. L. & Nolan, P. E. in *The Organisation and Expression of the Mitochondrial Genome* (eds Kroon, A. M. & Saccone, C.) 97-102 (Elsevier, Amsterdam, 1980).
10. Vierny, C., Keller, A. M., Begel, O. & Belcour, L. *Nature* **297**, 157-159 (1982).
11. Koll, F., Begel, O., Keller, A. M., Vierny, C. & Belcour, L. *Curr. Genet.* **8**, 127-134 (1984).
12. Osiewacz, H. D. & Esser, K. *Curr. Genet.* **8**, 299-305 (1984).
13. Jamet-Vierny, C., Begel, O. & Belcour, L. *Eur. J. Biochem.* **143**, 389-394 (1984).
14. Belcour, L. & Vierny, C. *EMBO J.* **5**, 609-614 (1986).
15. Koll, F., Vierny, C. & Belcour, L. *Plasmid* **14**, 106-117 (1985).
16. Shmookler Reis, R. J., Lumpking Jr, C. K., McGill, J. R., Riabowol, K. T. & Goldstein, S. *Nature* **316**, 167 (1985).
17. Barbacid, M. *Science* **225**, 670 (1984).
18. Marcou, D. *Ann. sci. natn. bot.* **11**, 653-764 (1961).
19. Debuchy, R. & Brygoo, Y. *EMBO J.* **4**, 3553-3556 (1985).
20. Rigby, P. W., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237-251 (1977).
21. Maniatis, T., Fritsch, E. & Sambrook, J. *Molecule Cloning; a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
22. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1985).

Thermal erosion by komatiites at Kambalda

GROVES *et al.*[1] have added a further twist to the continuing debate on thermal erosion by komatiite magma. This concept was introduced by Huppert *et al.*[2,3] who argue that hot, low-viscosity komatiite lava flows turbulently and transmits heat efficiently to floor rocks. When the flow rate is high enough and flowage continues long enough, floor rocks melt and are assimilated by the komatiite.

Groves *et al.*[1] present geological observations supporting thermal erosion, but claim the process operates only when a sulphide layer lies at the base of a flow between komatiite lava and floor rocks. They suggest that the sulphide transports heat from the komatiite to the substrate because of its high thermal conductivity, thereby promoting melting. I believe this concept is implausible on fluid dynamic grounds, and that some of the geological observations of Groves *et al.*[1] are open to alternative interpretation.

Figure 1*a* shows the temperature distribution in turbulently-flowing komatiite lava. The temperature remains constant nearly to the lower contact with the country rock, and decreases only in a thin boundary layer. Heat from the 1600 °C komatiite is transported directly to the floor rocks, causing them to melt. The effect of interposing a layer of any other material between the komatiite and the country rock substrate (Fig. 1*b*), even a layer of highly conductive sulphide, is to decrease the rate of heat transport, and to diminish the likelihood of melting. Using parameters listed by Huppert and Sparks[3], and assuming a conductivity for molten sulphide of 10 $Wm^{-1} K^{-1}$, I calculate that a 1 m thick sulphide layer decreases heat flux by around two orders of magnitude, if heat transport is solely by convection. Since the sulphide layer is heated from above, thermal convection is unimportant; if flowing komatiite stirs the sulphide, its insulating effect is diminished, but not eliminated.

The apparent association between erosion features and sulphide[1] may result because dense, highly fluid sulphide would naturally accumulate in depressions in the base of the flow, including any formed by thermal erosion.

The komatiite flows envisaged by Groves *et al.*[1] are not lethargic, merely capricious. These authors present convincing evidence that the second komatiite flow in the Kambalda sequence eroded 1 m from the top of the first komatiite flow. For komatiite to be eroded, it must be heated to 1400–1500 °C. Only then is it sufficiently molten to be removed by overriding lava. A lava flow capable of eroding komatiite would erode basalt (which melts

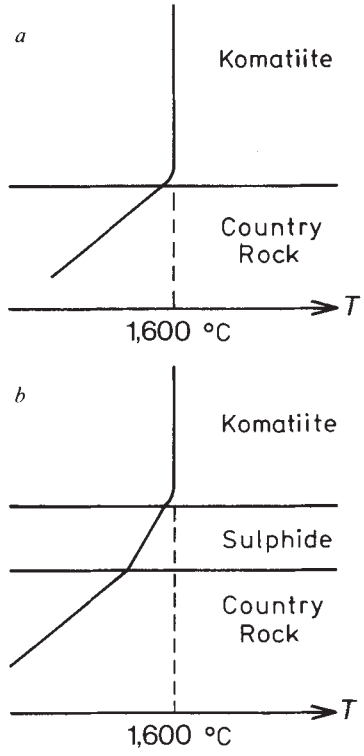


Fig. 1 *a*, Temperature (T) distribution in a komatiite flow and underlying rock and *b*, the insulating effect of an intervening sulphide layer.

at 1200 °C) three times faster, and sediment (which melts at 900 °C) seven times faster[2]. The lowermost komatiite flow at Kambalda, which is thicker and probably more magnesian than its successor[4], should have been capable of producing deep troughs in its basaltic and sedimentary substrate, not merely "second-order irregularities"[1].

Most of the geological features that lead Groves *et al.*[1] to reject the idea of extensive thermal erosion can be explained if we remember that volcanic eruptions are not continuous and that lava does not always follow the same course, but meanders, dams up, and spills out in different directions as the eruption rate fluctuates. Normally eruption rate climbs to a maximum, then declines. At some stage during the decline, the komatiite becomes incapable of erosion. Note also that erosion is not instantaneous but proceeds at a finite rate. Preservation of sediment beneath a komatiite flow may therefore indicate that erosion ceased because the flow rate declined, or the course of the komatiite lava changed. Erupting komatiite initially follows pre-existing valleys, and some primary topographic control is likely. The komatiite flow deepens these valleys, and flowage is concentrated in central feeder channels; a contrast in texture and composition between feeders and flanking units is to be expected. When komatiite encounters a permeable, water-saturated horizon such as flow-top or pillow breccia, hydrothermal circulation removes heat from the komatiite, and the erosion rate decreases. There is an enhanced probability of erosion ceasing when the komatiite reaches a permeable water saturated horizon, and this may appear to be caused by stratigraphic control.

Thermal erosion by komatiite flows is a new concept which may explain some of the geological and chemical features of komatiites that have puzzled us in the past. The task now is to return to mines like Kambalda, or to other areas where komatiites are well exposed to collect new petrologic and chemical data, and to use this information to test the concept. Groves *et al.*[1] have done this in their investigation of interspinifex sulphides, but other aspects of Kambalda geology seem open to alternative interpretation.

N. T. ARNDT
Max-Planck Institute für Chemie,
Postfach 3060,
D-6500 Mainz, FGR

1. Groves, D. I., Korkiakoski, E. A., McNaughton, N. J., Lesher, C. J. & Cowden, A. *Nature* **319**, 136–139 (1986).
2. Huppert, H. E., Sparks, R. S. J., Turner, J. S. & Arndt, N. T. *Nature* **309**, 19–23 (1984).
3. Huppert, H. E. & Sparks, R. S. J. *J. Petrol.* **26**, 694–725 (1986).
4. Lesher, C. M. thesis, Univ. Western Australia (1983).

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

Yes! I would like to know more about UMI Article Clearinghouse. I am interested in electronic ordering through the following system(s):
☐ DIALOG/Dialorder ☐ ITT Dialcom
☐ OnTyme ☐ OCLC ILL Subsystem
☐ Other (please specify)_____
☐ I am interested in sending my order by mail.
☐ Please send me your current catalog and user instructions for the system(s) I checked above.
Name_____
Title_____
Institution/Company_____
Department_____
Address_____
City_____ State_____ Zip_____
Phone (____)_____

UMI Article Clearinghouse

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106